

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
12 August 2004 (12.08.2004)

PCT

(10) International Publication Number  
**WO 2004/067743 A1**

(51) International Patent Classification<sup>7</sup>: C12N 15/11,  
15/63, C07H 21/04, A01H 5/00

(21) International Application Number:  
PCT/US2003/041269

K. [US/US]; 111 Pecan Meadow Drive, Baton Rouge, LA 70810 (US). FIORETTI, William, C. [US/US]; 2225 Lak-  
eridge Road Drive, Grapevine, TX 76051 (US). CADD,  
Gary, G. [US/US]; 501 Turner Road, Apartment 1111,  
Grapevine, TX 76051 (US).

(22) International Filing Date:  
24 December 2003 (24.12.2003)

(74) Agent: PRATT, John, S.; Kilpatrick Stockton LLP, Suite  
2800, 1100 Peachtree Street, Atlanta, GA 30309 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/441,377	21 January 2003 (21.01.2003)	US
60/441,381	21 January 2003 (21.01.2003)	US
60/441,392	21 January 2003 (21.01.2003)	US
607441,405	21 January 2003 (21.01.2003)	US
60/441,447	21 January 2003 (21.01.2003)	US
60/441,502	21 January 2003 (21.01.2003)	US
10/609,019	26 June 2003 (26.06.2003)	US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR,  
CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD,  
GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR,  
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN,  
MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU,  
SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA,  
UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (BW, GH,  
GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,  
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE,  
SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA,  
GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicants (*for all designated States except US*): TRANS-  
GENRX, INC. [US/US]; 1755 Wittington Place Drive,  
Suite 300, Dallas, TX 75234 (US). THE BOARD OF  
SUPERVISORS OF LOUISIANA STATE UNIVER-  
SITY AND AGRICULTURAL AND MECHANICAL  
COLLEGE [US/US]; P.O. Box 25055, Baton Rouge, LA  
70894 (US).

Published:

— with international search report

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): COOPER, Richard,

*For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.*

(54) Title: VACCINE PRODUCTION USING TRANSPOSON BASED VECTORS

(57) Abstract: The present invention provides novel methods of making proteins, multimeric proteins and antibodies using transpo-  
son-based vectors in transgenic individuals that are desirable in producing vaccines. The present invention further provides a method  
for providing protection against disease by vaccinating an individual with the proteins, multimeric proteins and/or antibodies made  
in the transgenic individuals of the present invention.

BEST AVAILABLE COPY

WO 2004/067743 A1



5

## 10 VACCINE PRODUCTION USING TRANSPOSON BASED VECTORS

The U.S. Government has certain rights in this invention. The development of this invention was partially funded by the United States Government under a HATCH  
15 grant from the United States Department of Agriculture, partially funded by the United States Government with Formula 1433 funds from the United States Department of Agriculture and partially funded by the United States Government under contract DAAD 19-02016 awarded by the Army.

## 20 FIELD OF THE INVENTION

The present invention relates generally to production of proteins, polypeptides and peptides in a transgenic individual, wherein genes encoding the proteins, polypeptides and peptides are operably-linked to signal sequences, or portions of signal sequences. The proteins, polypeptides and peptides are useful as vaccines.

25

## BACKGROUND OF THE INVENTION

Diseases caused by numerous biological organisms result in widespread morbidity and mortality throughout the world. Infectious disease is rampant in many locations, and unfortunately, numerous individuals suffer from more than one disease.  
30 Immunization of animals and humans with compositions comprising vaccines capable of generating an immune response is one of the only means of preventing or ameliorating the severity of disease.

Numerous diseases threaten not only humans but also animals. Influenza epidemics kill millions each year. Supply of influenza vaccine is often inadequate for

the needs of numerous individuals. Further, the influenza vaccine is often designed on the best guesses of infectious disease experts concerning the expected antigens and strains for any given year. Different governments and the Centers for Disease Control face a major challenge in redesigning a vaccine and rapidly producing it to supply those in need.

Recent bioterrorism, for example the anthrax attacks, demonstrates the need for a stockpile of antidote. Bioterrorism against animals may also threaten the food supplies of nations. Engineered organisms provide a particular challenge for nations to rapidly adjust their vaccine production to attack newly discovered epitopes in these organisms.

Recent events have demonstrated the vulnerability of the U.S. to attack with biowarfare agents. For many of the agents listed as select agents by the Centers for Disease Control, vaccines are not readily available for humans or domestic livestock. Despite intense research with many of these organisms, the potential for a vaccine in the immediate future is not likely. In addition, treatment with antibiotics may or may not be effective, especially if a bacterium has been engineered to be multi-drug resistant.

Accordingly what is needed in a method for rapid and economical production of proteins, polypeptides, peptides and mixtures thereof that may be used as vaccines to prevent or ameliorate disease.

## SUMMARY OF THE INVENTION

The present invention provides a new, effective and efficient method of producing proteins, polypeptides and peptides (hereinafter generally "proteins") in a transgenic individual that may be used as vaccines. The present invention provides the capability to transform the genetic machinery of individuals in order to produce vaccines. Such transformation may be germ-line transformation and/or non germ-line transformation. The present invention facilitates production of one or more protein, polypeptide and peptide for use in vaccines.

This invention provides polynucleotide cassettes containing at least one gene of interest and one or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro nucleotide sequence. Each of the genes of interest encodes a protein that may be used in the vaccine. One discovery of the present



invention is the use of pro portions of prepro signal sequences to facilitate appropriate processing, expression, and/or formation of proteins in an individual. Preferred pro polynucleotides are provided in SEQ ID NO:1 and SEQ ID NO:2.

Another discovery of the present invention is that cecropin prepro sequences  
 5 facilitate appropriate processing, expression, and/or formation of desired proteins in an individual. Accordingly, the present invention includes polynucleotide cassettes containing at least one gene of interest operably-linked to a cecropin prepro sequence. Preferred cecropin prepro polynucleotides are provided in SEQ ID NO:3 and SEQ ID NO:4. The present invention also includes polynucleotide cassettes containing at least  
 10 one gene of interest operably linked to a cecropin prepro polynucleotide, wherein pro sequences are located between the at least one gene of interest when more than one gene of interest is present.

A non-limiting list of the diseases and organisms that may be targets of the vaccines of the present invention follows: Anthrax (*Bacillus anthracis*), *Bacillus*  
 15 *anthracis* (anthrax), Botulism (*Clostridium botulinum* toxin), *Brucella* species (brucellosis), Brucellosis (*Brucella* species), *Burkholderia mallei* (glanders), *Burkholderia pseudomallei* (melioidosis), *Chlamydia psittaci* (psittacosis), Cholera (*Vibrio cholerae*), *Clostridium botulinum* toxin (botulism), *Clostridium perfringens* (Epsilon toxin), *Coxiella burnetii* (Q fever), *E. coli* O157:H7 (*Escherichia coli*),  
 20 Emerging infectious diseases such as Nipah virus and hantavirus, Epsilon toxin of *Clostridium perfringens*, *Escherichia coli* O157:H7 (*E. coli*), Food safety threats (e.g., *Salmonella* species, *Escherichia coli* O157:H7, *Shigella*), *Francisella tularensis* (tularemia), Glanders (*Burkholderia mallei*), Melioidosis (*Burkholderia pseudomallei*), Plague (*Yersinia pestis*), Psittacosis (*Chlamydia psittaci*), Q fever  
 25 (*Coxiella burnetii*), Ricin toxin from *Ricinus communis* (castor beans), *Rickettsia prowazekii* (typhus fever), *Salmonella* species (salmonellosis), *Salmonella* Typhi (typhoid fever), Salmonellosis (*Salmonella* species), *Shigella* (shigellosis), Shigellosis (*Shigella*), Smallpox (variola major), Staphylococcal enterotoxin B, Tularemia (*Francisella tularensis*), Typhoid fever (*Salmonella* Typhi), Typhus fever (*Rickettsia*  
 30 *prowazekii*), Variola major (smallpox), *Vibrio cholerae* (cholera), Viral encephalitis (alphaviruses [e.g., Venezuelan equine encephalitis, eastern equine encephalitis, western equine encephalitis]), Viral hemorrhagic fevers (filoviruses [e.g., Ebola, Marburg] and arenaviruses [e.g., Lassa, Machupo]), Water safety threats (e.g., *Vibrio*

*cholerae*, *Cryptosporidium parvum*), and *Yersinia pestis* (plague). Other diseases and organisms that may be targets of the vaccines of the present invention are described in Harrison's Principles of Internal Medicine, (Fauci et al. eds, 14<sup>th</sup> edition, McGraw Hill, New York, USA (1998)).

5           These polynucleotide cassettes are administered to an individual for expression of polypeptide sequences and the formation of a desired protein or multimeric protein for use as a vaccine. Preferably, the individual is an animal from which the multimeric protein can be harvested. In one embodiment, preferred animals are egg-laying or milk-producing animals.

10           In one embodiment, the egg-laying transgenic animal is an avian. The method of the present invention may be used in avians including Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Passeriformes, Coraciiformes, Ralliformes, Cuculiformes, Columbiformes, Galliformes, Anseriformes, and Herodiones. Preferably, the egg-laying transgenic animal is a poultry bird. More preferably, the  
15           bird is a chicken, turkey, duck, goose or quail. Another preferred bird is a ratite, such as, an emu, an ostrich, a rhea, or a cassowary. Other preferred birds are partridge, pheasant, kiwi, parrot, parakeet, macaw, falcon, eagle, hawk, pigeon, cockatoo, song birds, jay bird, blackbird, finch, warbler, canary, toucan, mynah, or sparrow.

          In some embodiments, the polynucleotide cassettes are located within  
20           transposon-based vectors that allow for incorporation of the cassettes into the DNA of the individual. The transposon-based vectors of the present invention include a transposase, operably-linked to a first promoter, and a coding sequence for a protein or peptide of interest operably-linked to a second promoter, wherein the coding sequence for the protein or peptide of interest and its operably-linked promoter are  
25           flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes the following characteristics: a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the  
30           nucleotide at the third base position of each codon is changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of

the transposase gene. In some embodiments, the effective polyA sequence is an avian optimized polyA sequence.

In one embodiment, the transposon-based vector comprises an avian optimized polyA sequence and does not comprise a modified Kozak sequence comprising  
5 ACCATG (SEQ ID NO:5). One example of such a transposon-based vector is the pTnMCS vector (SEQ ID NO:56). In another embodiment the transposon-based vector comprises a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino  
10 acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. One example of such a transposon-based vector is  
15 the pTnMod vector (SEQ ID NO:57).

Accordingly, it is an object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines.

It is another object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines, wherein the proteins are antibodies.

20 It is another object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines, wherein the proteins are multimeric proteins.

It is another object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines, wherein the proteins are monoclonal  
25 antibodies.

Yet another object of the present invention is to vaccinate individuals, particularly animals and humans, with the vaccines of the present invention in order to prevent or ameliorate disease.

Yet another object of the present invention is to vaccinate individuals,  
30 particularly animals and humans, with a vaccine comprising a combination of the proteins or multimeric proteins of the present invention, together with antibodies of the present invention, in order to prevent or ameliorate disease.

Another object of the present invention is to provide proteins, multimeric proteins and antibodies that are useful in the preparation of a medicament that may be used as a vaccine.

An advantage of the present invention is that the proteins are produced more efficiently, in greater quantities and less expensively than taught by prior art methods.

These and other objects, features and advantages of the present invention will become apparent after a review of the following detailed description of the disclosed embodiments and claims.

## BRIEF DESCRIPTION OF THE FIGURES

Figure 1 depicts schematically a polynucleotide cassette containing two genes of interest operably-linked to two pro polynucleotides, wherein the first pro polynucleotide is a part of a prepro polynucleotide. "Prom" indicates promoter.

Figure 2 depicts schematically a polynucleotide cassette containing polynucleotides encoding for a heavy chain and a light chain of an antibody. "Oval prom" indicates an ovalbumin promoter. The polynucleotide cassette contains pro and prepro sequences and is flanked by insertion sequences (IS) recognized by a transposase.

Figure 3 depicts schematically a polynucleotide cassette containing a cecropin prepro sequence operably-linked to two genes of interest. Between the genes of interest resides a cleavage site indicated by "CS."

Figure 4 depicts schematically a polynucleotide cassette containing two genes of interest, a promoter (prom), a signal sequence (SS) and a cleavage site (CS). The polynucleotide cassette is flanked by insertion sequences (IS) recognized by a transposase.

Figure 5 is a picture of a gel showing partially purified egg white derived from a transgenic avian run under reducing and non-reducing conditions.

## DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a new, effective and efficient method of producing proteins in an individual that may be used as vaccines. Such proteins include polypeptides, peptides and mixtures thereof and are generally called "proteins" throughout the present application. One or more proteins may be made

with the present invention. Antibodies may be made with the present invention and used as vaccines, alone or in conjunction with other proteins and multimeric proteins. Multimeric proteins may be produced with the present invention and include at least two proteins, wherein the at least two proteins may be the same or different. A  
5 multimeric protein includes associated multimeric proteins (two or more associated polypeptides) and multivalent multimeric proteins (a single polypeptide encoded by more than one gene of interest). Expression and/or formation of the multimeric protein in the individual is achieved by administering a polynucleotide cassette containing the gene or genes of interest to the individual. The polynucleotide cassette  
10 may additionally contain one or more pro sequences, prepro sequences, cecropin prepro sequences, and/or cleavage site sequences.

One or more genes of interest for incorporation into the polynucleotide cassette are chosen based on the genes that generally code for epitopes of organisms that are believed to be antigenic. Such coding sequences are generally known to one  
15 of ordinary skill in the art and can be obtained through GenBank and the published scientific and patent literature. It is within the scope of the present invention to search the genome of organisms to identify coding regions for particular proteins, polypeptides, peptides or epitopes that could serve as epitopes for vaccines. A protein may be made for use in a vaccine against any desired disease or disease producing  
20 organism using the method of the present invention. A non-limiting list of the diseases and organisms that may be targets of the vaccines of the present invention follows: Anthrax (*Bacillus anthracis*), *Bacillus anthracis* (anthrax), Botulism (*Clostridium botulinum* toxin), *Brucella* species (brucellosis), Brucellosis (*Brucella* species), *Burkholderia mallei* (glanders), *Burkholderia pseudomallei* (melioidosis),  
25 *Chlamydia psittaci* (psittacosis), Cholera (*Vibrio cholerae*), *Clostridium botulinum* toxin (botulism), *Clostridium perfringens* (Epsilon toxin), *Coxiella burnetii* (Q fever), *E. coli* O157:H7 (*Escherichia coli*), Emerging infectious diseases such as Nipah virus and hantavirus, Epsilon toxin of *Clostridium perfringens*, *Escherichia coli* O157:H7 (*E. coli*), Food safety threats (e.g., *Salmonella* species, *Escherichia coli* O157:H7,  
30 *Shigella*), *Francisella tularensis* (tularemia), Glanders (*Burkholderia mallei*), Melioidosis (*Burkholderia pseudomallei*), Plague (*Yersinia pestis*), Psittacosis (*Chlamydia psittaci*), Q fever (*Coxiella burnetii*), Ricin toxin from *Ricinus communis* (castor beans), *Rickettsia prowazekii* (typhus fever), *Salmonella* species

(salmonellosis), *Salmonella* Typhi (typhoid fever), Salmonellosis (*Salmonella* species), *Shigella* (shigellosis), Shigellosis (*Shigella*), Smallpox (variola major), Staphylococcal enterotoxin B, Tularemia (*Francisella tularensis*), Typhoid fever (*Salmonella* Typhi), Typhus fever (*Rickettsia prowazekii*), Variola major (smallpox),  
5 *Vibrio cholerae* (cholera), Viral encephalitis (alphaviruses [e.g., Venezuelan equine encephalitis, eastern equine encephalitis, western equine encephalitis]), Viral hemorrhagic fevers (filoviruses [e.g., Ebola, Marburg] and arenaviruses [e.g., Lassa, Machupo]), Water safety threats (e.g., *Vibrio cholerae*, *Cryptosporidium parvum*), and *Yersinia pestis* (plague). Other diseases and organisms that may be targets of the  
10 vaccines of the present invention are described in Harrison's Principles of Internal Medicine, (Fauci et al. eds, 14<sup>th</sup> edition, McGraw Hill, New York, USA (1998)).

This invention provides polynucleotide cassettes containing at least one gene of interest and one or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro nucleotide sequence. When one gene of interest is  
15 present in the polynucleotide cassette, a single protein is made, as opposed to a multimeric protein as described below. The present invention provides polynucleotide cassettes containing two or more genes of interest and two or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro nucleotide sequence. These polynucleotide cassettes are useful for making  
20 multimeric proteins. Each of the genes of interest encodes a polypeptide that forms a part of the multimeric protein. These polynucleotide cassettes are administered to an individual for expression of the polypeptide sequences and expression and/or formation of the multimeric protein. Preferably, the individual is an animal from which the proteins or multimeric protein can be harvested. Preferred animals are egg-  
25 laying or milk-producing animals. In some embodiments, the polynucleotide cassettes are located within transposon-based vectors that allow for incorporation of the cassettes into the DNA of the individual.

The pro polynucleotide sequences operably-linked to the genes of interest include pro portions of prepro polynucleotide sequences commonly associated with  
30 polynucleotides encoding proteins secreted from a cell in nature. It may be that the pre polynucleotide sequence functions to direct the resultant protein into the endoplasmic reticulum and the pro sequence is cleaved within the endoplasmic reticulum or Golgi complex of a cell containing the protein. While prepro

polynucleotide sequences are associated with secreted polypeptides in nature, one discovery of the present invention is the use of pro portions of the prepro signal sequences to facilitate appropriate processing, expression, and/or formation of proteins and multimeric proteins, and more particularly, associated multimeric proteins. In the present invention, each gene of interest is operably-linked with a pro polynucleotide sequence. Figure 1 shows schematically one polynucleotide cassette containing two genes of interest, wherein each gene of interest is operably-linked to a pro polynucleotide sequence. The first gene of interest is operably-linked to a pro polynucleotide sequence that is part of a prepro polynucleotide sequence, while the second gene of interest is operably-linked to a pro polynucleotide sequence that is not part of a prepro polynucleotide sequence, but may have been derived from a prepro polynucleotide sequence. Accordingly, the term "pro sequence" encompasses a pro sequence that is part of a prepro sequence and a pro sequence that is not part of a prepro sequence, but may have been derived from a prepro sequence. In preferred embodiments, the most 5' pro polynucleotide sequence in the polynucleotide cassette is a part of a prepro polynucleotide sequence.

Several examples of prepro polynucleotides from which a pro polynucleotide can be derived or be a part of are a cecropin prepro, lysozyme prepro, ovomucin prepro, ovotransferrin prepro, a signal peptide for tumor necrosis factor receptor (SEQ ID NO:6), a signal peptide encoded by a polynucleotide sequence provided in one of SEQ ID NOs:7-54 and a signal peptide provide in SEQ ID NO:55. The prepro or pro polynucleotide can be a cecropin prepro or pro polynucleotide selected from the group consisting of cecropin A1, cecropin A2, cecropin B, cecropin C, cecropin D, cecropin E and cecropin F. In a preferred embodiment, the pro polynucleotide is a cecropin B pro polynucleotide having a sequence shown in SEQ ID NO:1 or SEQ ID NO:2. A preferred prepro polynucleotide is a cecropin B polynucleotide having a sequence shown in SEQ ID NO:3 or SEQ ID NO:4.

Figure 1 provides one embodiment of the invention wherein the polynucleotide cassette includes two genes of interest and two pro polynucleotide sequences arranged in the following order: a prepro polynucleotide, a first gene of interest, a pro polynucleotide, and a second gene of interest. Preferably, the sequences are arranged in the aforementioned order beginning at a 5' end of the polynucleotide cassette. Figure 2 provides a more specific embodiment of the present

invention wherein the first and second genes of interest are polynucleotides encoding antibody heavy and light chains. However, the invention includes polynucleotide cassettes containing at least two genes of interest. Each of the genes of interest is operably-linked to a pro polynucleotide. Each of these pro polynucleotides can be the same, or each can be different. In one embodiment, all of the pro polynucleotides in the polynucleotide cassette are the same and are cecropin pro polynucleotides. The most 5' cecropin pro polynucleotide is preferably a part of a cecropin prepro polynucleotide sequence as shown in Figure 3.

The polynucleotide cassettes of the present invention may be administered to an individual for production of a protein or a multimeric protein in that individual. The discussion that follows is intended to encompass methods for making one protein or a multimeric protein. Accordingly, the present invention includes a method of producing a protein in an individual comprising administering to the individual a polynucleotide cassette comprising one gene of interest encoding the protein, wherein the gene of interest is operably-linked to a pro polynucleotide sequence. The present invention includes a method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide cassette comprising at least two genes of interest, each encoding a part of the multimeric protein, wherein each gene of interest is operably-linked to a pro polynucleotide sequence. The present invention also includes a method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide cassette comprising a cecropin prepro sequence operably-linked to two or more genes of interest, each gene of interest encoding a part of the multimeric protein. This second method does not require the linking of pro polynucleotides to each gene of interest since the use of a cecropin prepro sequence itself in a polynucleotide cassette facilitates processing, expression, and/or formation of multimeric proteins. Polynucleotide cassettes containing the cecropin prepro polynucleotide can contain one or at least two genes of interest. Preferably, the cecropin prepro polynucleotide is located 5' of the one or more genes of interest in the polynucleotide cassette. One exemplary polynucleotide cassette is shown in Figure 3. In a preferred embodiment, the prepro sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4. As shown in Figure 3, the polynucleotide cassettes containing a cecropin prepro polynucleotide preferably contain a cleavage site between each of two genes of interest. Such cleavage site(s)



may be nucleotides encoding any cleavage sites including, but not limited to, an enzymatic cleavage site, a pro polynucleotide, and a photolabile cleavage site, a chemical cleavage site, and a self-splicing cleavage site (i.e., intein). Cleavage sites are discussed in more detail below.

5       The polynucleotide cassettes of the present invention are particularly suited for production of proteins and/or multimeric proteins in an individual. Individuals include both humans and animals. Preferred animals are egg-laying animals and milk-producing animals. As used herein, the term "egg-laying animal" includes all amniotes such as birds, turtles, lizards and monotremes. Monotremes are egg-laying  
10       mammals and include the platypus and echidna. The term "bird" or "fowl," as used herein, is defined as a member of the Aves class of animals which are characterized as warm-blooded, egg-laying vertebrates primarily adapted for flying. Avians include, without limitation, Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Passeriformes, Coraciformes, Ralliformes, Cuculiformes, Columbiformes,  
15       Galliformes, Anseriformes, and Herodiones. The term "Ratite," as used herein, is defined as a group of flightless, mostly large, running birds comprising several orders and including the emus, ostriches, kiwis, and cassowaries. The term "Psittaciformes", as used herein, includes parrots and refers to a monofamilial order of birds that exhibit zygodactylism and have a strong hooked bill. A "parrot" is defined as any member of  
20       the avian family Psittacidae (the single family of the Psittaciformes), distinguished by the short, stout, strongly hooked beak. Preferred avians are poultry birds including chickens, quail, turkeys, geese and ducks. The term "chicken" as used herein denotes chickens used for table egg production, such as egg-type chickens, chickens reared for public meat consumption, or broilers, and chickens reared for both egg and meat  
25       production ("dual-purpose" chickens). The term "chicken" also denotes chickens produced by primary breeder companies, or chickens that are the parents, grandparents, great-grandparents, etc. of those chickens reared for public table egg, meat, or table egg and meat consumption.

      When the polynucleotide cassettes of the present invention are administered to  
30       an egg-laying or milk-producing animal, a transgenic animal containing a polynucleotide cassette is created and the animal produces a transgenic protein or multimeric protein. It is preferred that the resultant protein or multimeric protein is deposited in the egg or in the milk. Various different signal sequences and promoters

may be used to achieve deposition of the protein or multimeric protein in the egg or in the milk and these are described in more detail below. In order to achieve a transgenic animal containing a polynucleotide cassette of the present invention, the polynucleotide cassettes can be administered to the individual with, or contained in, any vector, as naked DNA, or in any delivery construct or solution commonly known to one of ordinary skill in the art. A preferred vector for incorporation of the polynucleotide cassettes into an individual is a transposon-based vector described below.

#### Definitions

It is to be understood that as used in the specification and in the claims, “a” or “an” can mean one or more, depending upon the context in which it is used. Thus, for example, reference to “a cell” can mean that at least one cell can be utilized.

The term protein includes polypeptide and peptide and these terms are generally encompassed under the term “protein” and are also used interchangeably herein. A multimeric protein includes more than one protein, polypeptide or peptide and is defined further below.

The term “antibody” is used interchangeably with the term “immunoglobulin” and is defined herein as a protein synthesized by an animal or a cell of the immune system in response to the presence of a foreign substance commonly referred to as an “antigen” or an “immunogen.”. The term antibody includes fragments of antibodies. Antibodies are characterized by specific affinity to a site on the antigen, wherein the site is referred to an “antigenic determinant” or an “epitope”. Antigens can be naturally occurring or artificially engineered. Preferred antigens are those from organisms against which a vaccine is desired. Antigens and suspected or known epitopes may be obtained from any organism, including but not limited to those recited herein.

Artificially engineered antigens include but are not limited to small molecules, such as small peptides, attached to haptens such as macromolecules, for example proteins, nucleic acids, or polysaccharides. Artificially designed or engineered variants of naturally occurring antibodies and artificially designed or engineered antibodies not occurring in nature are all included in the current definition. Such variants include conservatively substituted amino acids and other forms of substitution as described in the section concerning proteins and polypeptides.

The term "egg" is defined herein as including a large female sex cell enclosed in a porous, calcarous or leathery shell, produced by birds and reptiles. The term "ovum" is defined as a female gamete, and is also known as an egg. Therefore, egg production in all animals other than birds and reptiles, as used herein, is defined as the production and discharge of an ovum from an ovary, or "ovulation". Accordingly, it is to be understood that the term "egg" as used herein is defined as a large female sex cell enclosed in a porous, calcarous or leathery shell, when a bird or reptile produces it, or it is an ovum when it is produced by all other animals.

The term "gene" is defined herein to include a polynucleotide that includes a coding region for a protein, peptide or polypeptide, with or without intervening sequences such as introns.

The term "multimeric protein" is defined herein to include one or more polypeptides that are associated, or joined, by any means including disulfide bonds. An example of this type of multimeric protein is an antibody that contains both heavy and light chains that are associated by disulfide bonds. These multimeric proteins are referred to herein as "associated multimeric proteins." The term "multimeric protein" also includes a polypeptide that is encoded by more than one gene of interest, such multimeric protein containing desired amino acid sequences encoded by the more than one gene of interest. An example of this type of multimeric protein is a single polypeptide containing a heavy chain polypeptide (first polypeptide of interest) and a light chain polypeptide (second polypeptide of interest). In these embodiments, the different polypeptides of interest may be separated by other polypeptide sequences such as spacer polypeptides and cleavage site polypeptides. These types of multimeric proteins are referred to herein as "multivalent multimeric proteins."

The term "milk-producing animal" refers herein to mammals including, but not limited to, bovine, ovine, porcine, equine, and primate animals. Milk-producing animals include but are not limited to cows, llamas, camels, goats, reindeer, zebu, water buffalo, yak, horses, pigs, rabbits, non-human primates, and humans.

The term "transgenic animal" refers to an animal having at least a portion of the transposon-based vector DNA is incorporated into its DNA. While a transgenic animal includes an animal wherein the transposon-based vector DNA is incorporated into the germline DNA, a transgenic animal also includes an animal having DNA in one or more somatic cells that contain a portion of the transposon-based vector DNA

for any period of time. In a preferred embodiment, a portion of the transposon-based vector comprises a gene of interest. More preferably, the gene of interest is incorporated into the animal's DNA for a period of at least five days, more preferably the laying life of the animal, and most preferably the life of the animal. In a further  
5 preferred embodiment, the animal is an avian.

The term "vector" is used interchangeably with the terms "construct", "DNA construct" and "genetic construct" to denote synthetic nucleotide sequences used for manipulation of genetic material, including but not limited to cloning, subcloning, sequencing, or introduction of exogenous genetic material into cells, tissues or  
10 organisms, such as birds. It is understood by one skilled in the art that vectors may contain synthetic DNA sequences, naturally occurring DNA sequences, or both. The vectors of the present invention are transposon-based vectors as described herein.

When referring to two nucleotide sequences, one being a regulatory sequence, the term "operably-linked" is defined herein to mean that the two sequences are  
15 associated in a manner that allows the regulatory sequence to affect expression of the other nucleotide sequence. It is not required that the operably-linked sequences be directly adjacent to one another with no intervening sequence(s).

The term "regulatory sequence" is defined herein as including promoters, enhancers and other expression control elements such as polyadenylation sequences,  
20 matrix attachment sites, insulator regions for expression of multiple genes on a single construct, ribosome entry/attachment sites, introns that are able to enhance expression, and silencers.

#### Transposon-Based Vectors

While not wanting to be bound by the following statement, it is believed that  
25 the nature of the DNA construct is an important factor in successfully producing transgenic animals. The "standard" types of plasmid and viral vectors that have previously been almost universally used for transgenic work in all species, especially avians, have low efficiencies and may constitute a major reason for the low rates of transformation previously observed. The DNA (or RNA) constructs previously used  
30 often do not integrate into the host DNA, or integrate only at low frequencies. Other factors may have also played a part, such as poor entry of the vector into target cells. The present invention provides transposon-based vectors that can be administered to an animal that overcome the prior art problems relating to low transgene integration

frequencies. Two preferred transposon-based vectors of the present invention in which a transposase, gene of interest and other polynucleotide sequences may be introduced are termed pTnMCS (SEQ ID NO:56) and pTnMod (SEQ ID NO:57).

5 The transposon-based vectors of the present invention produce integration frequencies an order of magnitude greater than has been achieved with previous vectors. More specifically, intratesticular injections performed with a prior art transposon-based vector (described in U.S. Patent No. 5,719,055) resulted in 41% sperm positive roosters whereas intratesticular injections performed with the novel transposon-based vectors of the present invention resulted in 77% sperm positive  
10 roosters. Actual frequencies of integration were estimated by either or both comparative strength of the PCR signal from the sperm and histological evaluation of the testes and sperm by quantitative PCR.

The transposon-based vectors of the present invention include a transposase gene operably-linked to a first promoter, and a coding sequence for a desired protein  
15 or peptide operably-linked to a second promoter, wherein the coding sequence for the desired protein or peptide and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes one or more of the following characteristics: a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to  
20 enhance expression of the transposase; b) modifications of one or more of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the  
25 transposase to further enhance expression of the transposase gene. In one embodiment, the transposon-based vector comprises an avian optimized polyA sequence and does not comprise a modified Kozak sequence comprising ACCATG (SEQ ID NO:5). One example of such a transposon-based vector is the pTnMCS vector (SEQ ID NO:56). In another embodiment the transposon-based vector  
30 comprises a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without

changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. One example of such a transposon-based vector is the pTnMod  
5 vector (SEQ ID NO:57). The transposon-based vector may additionally or alternatively include one or more of the following Kozak sequences at the 3' end of any promoter, including the promoter operably-linked to the transposase: ACCATGG (SEQ ID NO:58), AAGATGT (SEQ ID NO:59), ACGATGA (SEQ ID NO:60), AAGATGG (SEQ ID NO:61), GACATGA (SEQ ID NO:62), ACCATGA (SEQ ID  
10 NO:63), and ACCATGA (SEQ ID NO:64), ACCATGT (SEQ ID NO:65).

#### Transposases and Insertion Sequences

In a further embodiment of the present invention, the transposase found in the transposase-based vector is an altered target site (ATS) transposase and the insertion sequences are those recognized by the ATS transposase. However, the transposase  
15 located in the transposase-based vectors is not limited to a modified ATS transposase and can be derived from any transposase. Transposases known in the prior art include those found in AC7, Tn5SEQ1, Tn916, Tn951, Tn1721, Tn 2410, Tn1681, Tn1, Tn2, Tn3, Tn4, Tn5, Tn6, Tn9, Tn10, Tn30, Tn101, Tn903, Tn501, Tn1000 ( $\gamma\delta$ ), Tn1681, Tn2901, AC transposons, Mp transposons, Spm transposons, En transposons, Dotted  
20 transposons, Mu transposons, Ds transposons, dSpm transposons and I transposons. According to the present invention, these transposases and their regulatory sequences are modified for improved functioning as follows: a) the addition one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the promoter operably-linked to the transposase; b) a change of one or more of the codons  
25 for the first several amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) the addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) the addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene.

30 Although not wanting to be bound by the following statement, it is believed that the modifications of the first several N-terminal codons of the transposase gene facilitate transcription of the transposase gene, in part, by increasing strand dissociation during transcription. It is preferable that one or more of between

approximately the first 1 to 20, more preferably 3 to 15, and most preferably between 4 to 12 N-terminal codons of the transposase are modified such that the third base of each codon is changed to an A or a T without changing the encoded amino acid. In one embodiment, the first ten N-terminal codons of the transposase gene are modified  
5 in this manner. It is also preferred that the transposase contain mutations that make it less specific for preferred insertion sites and thus increases the rate of transgene insertion as discussed in U.S. Patent No. 5,719,055.

In some embodiments, the transposon-based vectors are optimized for expression in a particular host by changing the methylation patterns of the vector  
10 DNA. For example, prokaryotic methylation may be reduced by using a methylation deficient organism for production of the transposon-based vector. The transposon-based vectors may also be methylated to resemble eukaryotic DNA for expression in a eukaryotic host.

Transposases and insertion sequences from other analogous eukaryotic  
15 transposon-based vectors that can also be modified and used are, for example, the *Drosophila* P element derived vectors disclosed in U.S. Patent No. 6,291,243; the *Drosophila* mariner element described in Sherman et al. (1998); or the sleeping beauty transposon. See also Hackett et al. (1999); D. Lampe et al., 1999. Proc. Natl. Acad. Sci. USA, 96:11428-11433; S. Fischer et al., 2001. Proc. Natl. Acad. Sci. USA,  
20 98:6759-6764; L. Zagoraiou et al., 2001. Proc. Natl. Acad. Sci. USA, 98:11474-11478; and D. Berg et al. (Eds.), *Mobile DNA*, Amer. Soc. Microbiol. (Washington, D.C., 1989). However, it should be noted that bacterial transposon-based elements are preferred, as there is less likelihood that a eukaryotic transposase in the recipient species will recognize prokaryotic insertion sequences bracketing the transgene.

Many transposases recognize different insertion sequences, and therefore, it is  
25 to be understood that a transposase-based vector will contain insertion sequences recognized by the particular transposase also found in the transposase-based vector. In a preferred embodiment of the invention, the insertion sequences have been shortened to about 70 base pairs in length as compared to those found in wild-type  
30 transposons that typically contain insertion sequences of well over 100 base pairs.

While the examples provided below incorporate a "cut and insert" Tn10 based vector that is destroyed following the insertion event, the present invention also encompasses the use of a "rolling replication" type transposon-based vector. Use of a

rolling replication type transposon allows multiple copies of the transposon/transgene to be made from a single transgene construct and the copies inserted. This type of transposon-based system thereby provides for insertion of multiple copies of a transgene into a single genome. A rolling replication type transposon-based vector may be preferred when the promoter operably-linked to gene of interest is endogenous to the host cell and present in a high copy number or highly expressed. However, use of a rolling replication system may require tight control to limit the insertion events to non-lethal levels. Tn1, Tn2, Tn3, Tn4, Tn5, Tn9, Tn21, Tn501, Tn551, Tn951, Tn1721, Tn2410 and Tn2603 are examples of a rolling replication type transposon, although Tn5 could be both a rolling replication and a cut and insert type transposon.

#### Stop Codons and PolyA Sequences

In one embodiment, the transposon-based vector contains two stop codons operably-linked to the transposase and/or to the gene of interest. In an alternate embodiment, one stop codon of UAA or UGA is operably linked to the transposase and/or to the gene of interest.

As used herein an "effective polyA sequence" refers to either a synthetic or non-synthetic sequence that contains multiple and sequential nucleotides containing an adenine base (an A polynucleotide string) and that increases expression of the gene to which it is operably-linked. A polyA sequence may be operably-linked to any gene in the transposon-based vector including, but not limited to, a transposase gene and a gene of interest. A preferred polyA sequence is optimized for use in the host animal or human. In one embodiment, the polyA sequence is optimized for use in an avian species and more specifically, a chicken. An avian optimized polyA sequence generally contains a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 75 base pairs that precede the A polynucleotide string and thereby separate the stop codon from the A polynucleotide string. In one embodiment of the present invention, the polyA sequence comprises a conalbumin polyA sequence as provided in SEQ ID NO:66 and as taken from GenBank accession # Y00407, base pairs 10651-11058. In another embodiment, the polyA sequence comprises a synthetic polynucleotide sequence shown in SEQ ID NO:67. In yet another embodiment, the polyA sequence comprises an avian optimized polyA sequence provided in SEQ ID NO:68. A chicken optimized



polyA sequence may also have a reduced amount of CT repeats as compared to a synthetic polyA sequence.

It is a surprising discovery of the present invention that such an avian optimized poly A sequence increases expression of a polynucleotide to which it is operably-linked in an avian as compared to a non-avian optimized polyA sequence. Accordingly, the present invention includes methods of or increasing incorporation of a gene of interest wherein the gene of interest resides in a transposon-based vector containing a transposase gene and wherein the transposase gene is operably linked to an avian optimized polyA sequence. The present invention also includes methods of increasing expression of a gene of interest in an avian that includes administering a gene of interest to the avian, wherein the gene of interest is operably-linked to an avian optimized polyA sequence. An avian optimized polyA nucleotide string is defined herein as a polynucleotide containing an A polynucleotide string and a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 75 base pairs that precede the A polynucleotide string. The present invention further provides transposon-based vectors containing a gene of interest or transposase gene operably linked to an avian optimized polyA sequence.

#### Promoters and Enhancers

The first promoter operably-linked to the transposase gene and the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. Constitutive promoters include, but are not limited to, immediate early cytomegalovirus (CMV) promoter, herpes simplex virus 1 (HSV1) immediate early promoter, SV40 promoter, lysozyme promoter, early and late CMV promoters, early and late HSV promoters,  $\beta$ -actin promoter, tubulin promoter, Rous-Sarcoma virus (RSV) promoter, and heat-shock protein (HSP) promoter. Inducible promoters include tissue-specific promoters, developmentally-regulated promoters and chemically inducible promoters. Examples of tissue-specific promoters include the glucose 6 phosphate (G6P) promoter, vitellogenin promoter, ovalbumin promoter, ovomucoid promoter, conalbumin promoter, ovotransferrin promoter, prolactin promoter, kidney uromodulin promoter, and placental lactogen promoter. In one embodiment, the vitellogenin promoter includes a polynucleotide sequence of SEQ ID NO:69. The G6P promoter sequence may be deduced from a rat G6P gene

untranslated upstream region provided in GenBank accession number U57552.1. Examples of developmentally-regulated promoters include the homeobox promoters and several hormone induced promoters. Examples of chemically inducible promoters include reproductive hormone induced promoters and antibiotic inducible promoters such as the tetracycline inducible promoter and the zinc-inducible metallothionine promoter.

Other inducible promoter systems include the Lac operator repressor system inducible by IPTG (isopropyl beta-D-thiogalactoside) (Cronin, A. et al. 2001. *Genes and Development*, v. 15), ecdysone-based inducible systems (Hoppe, U. C. et al. 2000. *Mol. Ther.* 1:159-164); estrogen-based inducible systems (Brasemann, S. et al. 1993. *Proc. Natl. Acad. Sci.* 90:1657-1661); progesterone-based inducible systems using a chimeric regulator, GLVP, which is a hybrid protein consisting of the GAL4 binding domain and the herpes simplex virus transcriptional activation domain, VP16, and a truncated form of the human progesterone receptor that retains the ability to bind ligand and can be turned on by RU486 (Wang, et al. 1994. *Proc. Natl. Acad. Sci.* 91:8180-8184); CID-based inducible systems using chemical inducers of dimerization (CIDs) to regulate gene expression, such as a system wherein rapamycin induces dimerization of the cellular proteins FKBP12 and FRAP (Belshaw, P. J. et al. 1996. *J. Chem. Biol.* 3:731-738; Fan, L. et al. 1999. *Hum. Gene Ther.* 10:2273-2285; Shariat, S.F. et al. 2001. *Cancer Res.* 61:2562-2571; Spencer, D.M. 1996. *Curr. Biol.* 6:839-847). Chemical substances that activate the chemically inducible promoters can be administered to the animal containing the transgene of interest via any method known to those of skill in the art.

Other examples of cell or tissue-specific and constitutive promoters include but are not limited to smooth-muscle SM22 promoter, including chimeric SM22alpha/telokin promoters (Hoggatt A.M. et al., 2002. *Circ Res.* 91(12):1151-9); ubiquitin C promoter (*Biochim Biophys Acta*, 2003. Jan. 3;1625(1):52-63); Hsf2 promoter; murine COMP (cartilage oligomeric matrix protein) promoter; early B cell-specific mb-1 promoter (Sigvardsson M., et al., 2002. *Mol. Cell Biol.* 22(24):8539-51); prostate specific antigen (PSA) promoter (Yoshimura I. et al., 2002, *J. Urol.* 168(6):2659-64); exorh promoter and pineal expression-promoting element (Asaoka Y., et al., 2002. *Proc. Natl. Acad. Sci.* 99(24):15456-61); neural and liver ceramidase gene promoters (Okino N. et al., 2002. *Biochem. Biophys. Res. Commun.*

299(1):160-6); PSP94 gene promoter/enhancer (Gabril M.Y. et al., 2002. Gene Ther. 9(23):1589-99); promoter of the human FAT/CD36 gene (Kuriki C., et al., 2002. Biol. Pharm. Bull. 25(11):1476-8); VL30 promoter (Staplin W.R. et al., 2002. Blood October 24, 2002); and IL-10 promoter (Brenner S., et al., 2002. J. Biol. Chem. 5 December 18, 2002).

Examples of avian promoters include, but are not limited to, promoters controlling expression of egg white proteins, such as ovalbumin, ovotransferrin (conalbumin), ovomucoid, lysozyme, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, ovostatin (ovomacroglobin), cystatin, avidin, thiamine-binding protein, glutamyl aminopeptidase minor glycoprotein 1, minor glycoprotein 2; and promoters controlling expression of egg-yolk proteins, such as vitellogenin, very low-density lipoproteins, low density lipoprotein, cobalamin-binding protein, riboflavin-binding protein, biotin-binding protein (Awade, 1996. Z. Lebensm. Unters. Forsch. 202:1-14). An advantage of using the vitellogenin promoter is that it is active during the egg-laying stage of an animal's life-cycle, which allows for the production of the protein of interest to be temporally connected to the import of the protein of interest into the egg yolk when the protein of interest is equipped with an appropriate targeting sequence. In some embodiments, the avian promoter is an oviduct-specific promoter. As used herein, the term "oviduct-specific promoter" includes, but is not limited to, ovalbumin, ovotransferrin (conalbumin), ovomucoid, lysozyme, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, and ovostatin (ovomacroglobin) promoters.

Liver-specific promoters of the present invention include, but are not limited to, the following promoters, vitellogenin promoter, G6P promoter, cholesterol-7-alpha-hydroxylase (CYP7A) promoter, phenylalanine hydroxylase (PAH) promoter, protein C gene promoter, insulin-like growth factor I (IGF-I) promoter, bilirubin UDP-glucuronosyltransferase promoter, aldolase B promoter, furin promoter, metallothioneine promoter, albumin promoter, and insulin promoter.

Also included in the present invention are promoters that can be used to target expression of a protein of interest into the milk of a milk-producing animal including, but not limited to,  $\beta$  lactoglobulin promoter, whey acidic protein promoter, lactalbumin promoter and casein promoter.

Promoters associated with cells of the immune system may also be used. Acute phase promoters such as interleukin (IL)-1 and IL-2 may be employed. Promoters for heavy and light chain Ig may also be employed. The promoters of the T cell receptor components CD4 and CD8, B cell promoters and the promoters of  
5 CR2 (complement receptor type 2) may also be employed. Immune system promoters are preferably used when the desired protein is an antibody protein.

Also included in this invention are modified promoters/enhancers wherein elements of a single promoter are duplicated, modified, or otherwise changed. In one embodiment, steroid hormone-binding domains of the ovalbumin promoter are moved  
10 from about -6.5 kb to within approximately the first 1000 base pairs of the gene of interest. Modifying an existing promoter with promoter/enhancer elements not found naturally in the promoter, as well as building an entirely synthetic promoter, or drawing promoter/enhancer elements from various genes together on a non-natural backbone, are all encompassed by the current invention.

15 Accordingly, it is to be understood that the promoters contained within the transposon-based vectors of the present invention may be entire promoter sequences or fragments of promoter sequences. For example, in one embodiment, the promoter operably linked to a gene of interest is an approximately 900 base pair fragment of a chicken ovalbumin promoter (SEQ ID NO:70). The constitutive and inducible  
20 promoters contained within the transposon-based vectors may also be modified by the addition of one or more modified Kozak sequences of ACCATG (SEQ ID NO:5).

As indicated above, the present invention includes transposon-based vectors containing one or more enhancers. These enhancers may or may not be operably-linked to their native promoter and may be located at any distance from their  
25 operably-linked promoter. A promoter operably-linked to an enhancer is referred to herein as an "enhanced promoter." The enhancers contained within the transposon-based vectors are preferably enhancers found in birds, and more preferably, an ovalbumin enhancer, but are not limited to these types of enhancers. In one embodiment, an approximately 675 base pair enhancer element of an ovalbumin  
30 promoter is cloned upstream of an ovalbumin promoter with 300 base pairs of spacer DNA separating the enhancer and promoter. In one embodiment, the enhancer used as a part of the present invention comprises base pairs 1-675 of a Chicken Ovalbumin

enhancer from GenBank accession #S82527.1. The polynucleotide sequence of this enhancer is provided in SEQ ID NO:71.

Also included in some of the transposon-based vectors of the present invention are cap sites and fragments of cap sites. In one embodiment, approximately 50 base  
5 pairs of a 5' untranslated region wherein the capsite resides are added on the 3' end of an enhanced promoter or promoter. An exemplary 5' untranslated region is provided in SEQ ID NO:72. A putative cap-site residing in this 5' untranslated region preferably comprises the polynucleotide sequence provided in SEQ ID NO:73.

In one embodiment of the present invention, the first promoter operably-linked  
10 to the transposase gene is a constitutive promoter and the second promoter operably-linked to the gene of interest is a tissue-specific promoter. In the second embodiment, use of the first constitutive promoter allows for constitutive activation of the transposase gene and incorporation of the gene of interest into virtually all cell types, including the germline of the recipient animal. Although the gene of interest is  
15 incorporated into the germline generally, the gene of interest is only expressed in a tissue-specific manner. A transposon-based vector having a constitutive promoter operably-linked to the transposase gene can be administered by any route, and in one embodiment, the vector is administered to an ovary or to an artery leading to the ovary. In another embodiment, the vector is administered into the lumen of the  
20 oviduct or into an artery supplying the oviduct.

It should be noted that cell- or tissue-specific expression as described herein does not require a complete absence of expression in cells or tissues other than the preferred cell or tissue. Instead, "cell-specific" or "tissue-specific" expression refers to a majority of the expression of a particular gene of interest in the preferred cell or  
25 tissue, respectively.

When incorporation of the gene of interest into the germline is not preferred, the first promoter operably-linked to the transposase gene can be a tissue-specific promoter. For example, transfection of a transposon-based vector containing a transposase gene operably-linked to an oviduct specific promoter such as the  
30 ovalbumin promoter provides for activation of the transposase gene and incorporation of the gene of interest in the cells of the oviduct but not into the germline and other cells generally. In this embodiment, the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. In a preferred

embodiment, both the first promoter and the second promoter are an ovalbumin promoter. In embodiments wherein tissue-specific expression or incorporation is desired, it is preferred that the transposon-based vector is administered directly to the tissue of interest or to an artery leading to the tissue of interest. In a preferred  
5 embodiment, the tissue of interest is the oviduct and administration is achieved by direct injection into the lumen of the oviduct or an artery leading to the oviduct. In a further preferred embodiment, administration is achieved by direct injection into the lumen of the magnum or the infundibulum of the oviduct.

Accordingly, cell specific promoters may be used to enhance transcription in  
10 selected tissues. In birds, for example, promoters that are found in cells of the fallopian tube, such as ovalbumin, conalbumin, ovomucoid and/or lysozyme, are used in the vectors to ensure transcription of the gene of interest in the epithelial cells and tubular gland cells of the fallopian tube, leading to synthesis of the desired protein encoded by the gene and deposition into the egg white. In mammals, promoters  
15 specific for the epithelial cells of the alveoli of the mammary gland, such as prolactin, insulin, beta lactoglobulin, whey acidic protein, lactalbumin, casein, and/or placental lactogen, are used in the design of vectors used for transfection of these cells for the production of desired proteins for deposition into the milk. In liver cells, the G6P promoter may be employed to drive transcription of the gene of interest for protein  
20 production. Proteins made in the liver of birds may be delivered to the egg yolk.

In order to achieve higher or more efficient expression of the transposase gene, the promoter and other regulatory sequences operably-linked to the transposase gene may be those derived from the host. These host specific regulatory sequences can be tissue specific as described above or can be of a constitutive nature. For  
25 example, an avian actin promoter and its associated polyA sequence can be operably-linked to a transposase in a transposase-based vector for transfection into an avian. Examples of other host specific promoters that could be operably-linked to the transposase include the myosin and DNA or RNA polymerase promoters.

#### Directing Sequences

30 In some embodiments of the present invention, the gene of interest is operably-linked to a directing sequence or a sequence that provides proper conformation to the desired protein encoded by the gene of interest. As used herein, the term "directing sequence" refers to both signal sequences and targeting sequences.

An egg directing sequence includes, but is not limited to, an ovomucoid signal sequence, an ovalbumin signal sequence, a cecropin pre pro signal sequence, and a vitellogenin targeting sequence. The term "signal sequence" refers to an amino acid sequence, or the polynucleotide sequence that encodes the amino acid sequence, a portion or the entirety of which directs the protein to which it is linked to the endoplasmic reticulum in a eukaryote, and more preferably the translocational pores in the endoplasmic reticulum, or the plasma membrane in a prokaryote, or mitochondria, such as for the purpose of gene therapy for mitochondrial diseases. Signal and targeting sequences can be used to direct a desired protein into, for example, the milk, when the transposon-based vectors are administered to a milk-producing animal.

Signal sequences can also be used to direct a desired protein into, for example, a secretory pathway for incorporation into the egg yolk or the egg white, when the transposon-based vectors are administered to a bird or other egg-laying animal. The present invention also includes a gene of interest operably-linked to a second gene containing a signal sequence. An example of such an embodiment is wherein the gene of interest is operably-linked to the ovalbumin gene that contains an ovalbumin signal sequence. Other signal sequences that can be included in the transposon-based vectors include, but are not limited to the ovotransferrin and lysozyme signal sequences. In one embodiment, the signal sequence is an ovalbumin signal sequence including a sequence shown in SEQ ID NO:74. In another embodiment, the signal sequence is a shortened ovalbumin signal sequence including a sequence shown in SEQ ID NO:75 or SEQ ID NO:76.

As also used herein, the term "targeting sequence" refers to an amino acid sequence, or the polynucleotide sequence encoding the amino acid sequence, which amino acid sequence is recognized by a receptor located on the exterior of a cell. Binding of the receptor to the targeting sequence results in uptake of the protein or peptide operably-linked to the targeting sequence by the cell. One example of a targeting sequence is a vitellogenin targeting sequence that is recognized by a vitellogenin receptor (or the low density lipoprotein receptor) on the exterior of an oocyte. In one embodiment, the vitellogenin targeting sequence includes the polynucleotide sequence of SEQ ID NO:77. In another embodiment, the vitellogenin targeting sequence includes all or part of the vitellogenin gene. Other targeting

sequences include VLDL and Apo E, which are also capable of binding the vitellogenin receptor. Since the ApoE protein is not endogenously expressed in birds, its presence may be used advantageously to identify birds carrying the transposon-based vectors of the present invention.

## 5 Genes of Interest

The genes of interest in the polynucleotide cassette can be any gene, and preferably are genes that encode proteins or portions of multimeric proteins that may be used in vaccines. Such genes may be coding regions for desired epitopes in infectious organisms, including but not limited to viral, bacterial, protozoal, fungal,  
10 and parasitic organisms. Such genes are known to one of ordinary skill in the art and may be obtained from GenBank and from scientific and patent literature. A gene of interest may contain modifications of the codons for the first several N-terminal amino acids of the gene of interest, wherein the third base of each codon is changed to an A or a T without changing the corresponding amino acid.

15 In one embodiment, the genes of interest are antibody genes or portions of antibody genes. Figure 2 shows a schematic drawing of a polynucleotide cassette containing an antibody heavy chain and an antibody light chain as two genes of interest. Antibodies used in or encoded by the polynucleotide cassettes of the present invention include, but are not limited to, IgG, IgM, IgA, IgD, IgE, IgY, lambda  
20 chains, kappa chains, bi-specific antibodies, and fragments thereof; scFv fragments, Fc fragments, and Fab fragments as well as dimeric, trimeric and oligomeric forms of antibody fragments. Desired antibodies include, but are not limited to, naturally occurring antibodies, human antibodies, humanized antibodies, autoantibodies and hybrid antibodies. Preferred antibodies include those useful as vaccines against  
25 selected diseases. Genes encoding modified versions of naturally occurring antibodies or fragments thereof and genes encoding artificially designed antibodies or fragments thereof may be incorporated into the transposon-based vectors of the present invention. Desired antibodies also include antibodies with the ability to bind specific ligands, for example, antibodies against proteins associated with diseases,  
30 particularly infectious diseases. Accordingly, the present invention encompasses a polynucleotide cassette as described herein containing one or more genes encoding a heavy immunoglobulin (Ig) chain and a light Ig chain.



Antibodies that may be produced using the present invention include, but are not limited to, antibodies for providing passive immunity to an animal or a human against an infectious disease or a toxic agent. The antibodies prepared using the methods of the present invention may also be designed to possess specific labels that may be detected through means known to one of ordinary skill in the art. For example, antibodies may be labeled with a fluorescent label attached that may be detected following exposure to specific wavelengths. Such labeled antibodies may be primary antibodies directed to a specific antigen, for example, rhodamine-labeled rabbit anti-growth hormone, or may be labeled secondary antibodies, such as fluorescein-labeled goat-anti chicken IgG. Such labeled antibodies are known to one of ordinary skill in the art. The antibodies may also be designed to possess specific sequences useful for purification through means known to one of ordinary skill in the art. Labels useful for attachment to antibodies are also known to one of ordinary skill in the art. Some of these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety. Proteins and multimeric proteins made according to the present invention may also be labeled with these methods. Antibodies, proteins and multimeric proteins produced with the present invention may be used as laboratory reagents for numerous applications including radioimmunoassay, western blots, dot blots, ELISA, immunoaffinity columns and other procedures requiring antibodies as known to one of ordinary skill in the art. Such antibodies include primary antibodies, secondary antibodies and tertiary antibodies, which may be labeled or unlabeled.

Additional antibodies that may be made with the practice of the present invention include, but are not limited to, primary antibodies, secondary antibodies, designer antibodies, anti-protein antibodies, anti-peptide antibodies, anti-DNA antibodies, anti-RNA antibodies, antibodies against non-natural antigens, anti-venom antibodies, antibodies directed against epitopes associated with infectious disease, including, anti-viral, anti-bacterial, anti-protozoal, anti-fungal, anti-parasitic, anti-receptor, anti-lipid, anti-phospholipid, anti-growth factor, anti-cytokine, anti-monokine, anti-idiotypic, and anti-accessory (presentation) protein antibodies. Antibodies made with the present invention, as well as light chains or heavy chains, may also be used to inhibit enzyme activity.

Antibodies that may be produced using the present invention include, but are not limited to, antibodies made against the following proteins: Bovine  $\gamma$ -Globulin, Serum; Bovine IgG, Plasma; Chicken  $\gamma$ -Globulin, Serum; Human  $\gamma$ -Globulin, Serum; Human IgA, Plasma; Human IgA<sub>1</sub>, Myeloma; Human IgA<sub>2</sub>, Myeloma; Human IgA<sub>2</sub>, Plasma; Human IgD, Plasma; Human IgE, Myeloma; Human IgG, Plasma; Human IgG, Fab Fragment, Plasma; Human IgG, F(ab')<sub>2</sub> Fragment, Plasma; Human IgG, Fc Fragment, Plasma; Human IgG<sub>1</sub>, Myeloma; Human IgG<sub>2</sub>, Myeloma; Human IgG<sub>3</sub>, Myeloma; Human IgG<sub>4</sub>, Myeloma; Human IgM, Myeloma; Human IgM, Plasma; Human Immunoglobulin, Light Chain  $\kappa$ , Urine; Human Immunoglobulin, Light Chains  $\kappa$  and  $\lambda$ , Plasma; Mouse  $\gamma$ -Globulin, Serum; Mouse IgG, Serum; Mouse IgM, Myeloma; Rabbit  $\gamma$ -Globulin, Serum; Rabbit IgG, Plasma; and Rat  $\gamma$ -Globulin, Serum. In one embodiment, the transposon-based vector comprises the coding sequence of light and heavy chains of a murine monoclonal antibody that shows specificity for human seminoprotein (GenBank Accession numbers AY129006 and AY129304 for the light and heavy chains, respectively).

A further non-limiting list of antibodies that recognize other antibodies and that may be produced using the present invention is as follows: Anti-Chicken IgG, heavy (H) & light (L) Chain Specific (Sheep); Anti-Goat  $\gamma$ -Globulin (Donkey); Anti-Goat IgG, Fc Fragment Specific (Rabbit); Anti-Guinea Pig  $\gamma$ -Globulin (Goat); Anti-Human Ig, Light Chain, Type  $\kappa$  Specific; Anti-Human Ig, Light Chain, Type  $\lambda$  Specific; Anti-Human IgA,  $\alpha$ -Chain Specific (Goat); Anti-Human IgA, Fab Fragment Specific; Anti-Human IgA, Fc Fragment Specific; Anti-Human IgA, Secretory; Anti-Human IgE,  $\epsilon$ -Chain Specific (Goat); Anti-Human IgE, Fc Fragment Specific; Anti-Human IgG, Fc Fragment Specific (Goat); Anti-Human IgG,  $\gamma$ -Chain Specific (Goat); Anti-Human IgG, Fc Fragment Specific; Anti-Human IgG, Fd Fragment Specific; Anti-Human IgG, H & L Chain Specific (Goat); Anti-Human IgG<sub>1</sub>, Fc Fragment Specific; Anti-Human IgG<sub>2</sub>, Fc Fragment Specific; Anti-Human IgG<sub>2</sub>, Fd Fragment Specific; Anti-Human IgG<sub>3</sub>, Hinge Specific; Anti-Human IgG<sub>4</sub>, Fc Fragment Specific; Anti-Human IgM, Fc Fragment Specific; Anti-Human IgM,  $\mu$ -Chain Specific; Anti-Mouse IgE,  $\epsilon$ -Chain Specific; Anti-Mouse  $\gamma$ -Globulin (Goat); Anti-Mouse IgG,  $\gamma$ -Chain Specific (Goat); Anti-Mouse IgG,  $\gamma$ -Chain Specific (Goat) F(ab')<sub>2</sub> Fragment; Anti-Mouse IgG, H & L Chain Specific (Goat); Anti-Mouse IgM,  $\mu$ -Chain Specific (Goat); Anti-Mouse IgM, H & L Chain Specific (Goat); Anti-Rabbit  $\gamma$ -Globulin

(Goat); Anti-Rabbit IgG, Fc Fragment Specific (Goat); Anti-Rabbit IgG, H & L Chain Specific (Goat); Anti-Rat  $\gamma$ -Globulin (Goat); Anti-Rat IgG, H & L Chain Specific; Anti-Rhesus Monkey  $\gamma$ -Globulin (Goat); and, Anti-Sheep IgG, H & L Chain Specific.

Antibodies may be made against peptides such as p146 (SEQ ID NO:78 amino acid sequence, SEQ ID NO:79, nucleotide sequence), and other lytic peptides or peptides associated with disease, particularly infectious disease.

The following is yet another non-limiting of antibodies that can be produced by the methods of present invention: abciximab (ReoPro), abciximab anti-platelet aggregation monoclonal antibody, anti-CD11a (hu1124), anti-CD18 antibody, anti-CD20 antibody, anti-cytomegalovirus (CMV) antibody, anti-digoxin antibody, anti-hepatitis B antibody, anti-HER-2 antibody, anti-idiotypic antibody to GD3 glycolipid, anti-IgE antibody, anti-IL-2R antibody, antimetastatic cancer antibody (mAb 17-1A), anti-rabies antibody, anti-respiratory syncytial virus (RSV) antibody, anti-Rh antibody, anti-TCR, anti-TNF antibody, anti-VEGF antibody and Fab fragment thereof, rattlesnake venom antibody, black widow spider venom antibody, coral snake venom antibody, antibody against very late antigen-4 (VLA-4), C225 humanized antibody to EGF receptor, chimeric (human & mouse) antibody against TNF $\alpha$ , antibody directed against GPIIb/IIIa receptor on human platelets, gamma globulin, anti-hepatitis B immunoglobulin, human anti-D immunoglobulin, human antibodies against *S aureus*, human tetanus immunoglobulin, humanized antibody against the epidermal growth receptor-2, humanized antibody against the  $\alpha$  subunit of the interleukin-2 receptor, humanized antibody CTLA4IG, humanized antibody to the IL-2 R  $\alpha$ -chain, humanized anti-CD40-ligand monoclonal antibody (5c8), humanized mAb against the epidermal growth receptor-2, humanized mAb to rous sarcoma virus, humanized recombinant antibody (IgG1k) against respiratory syncytial virus (RSV), lymphocyte immunoglobulin (anti-thymocyte antibody), lymphocyte immunoglobulin, mAb against factor VII, MDX-210 bi-specific antibody against HER-2, MDX-22, MDX-220 bi-specific antibody against TAG-72 on tumors, MDX-33 antibody to Fc $\gamma$ R1 receptor, MDX-447 bi-specific antibody against EGF receptor, MDX-447 bispecific humanized antibody to EGF receptor, MDX-RA immunotoxin (ricin A linked) antibody, Medi-507 antibody (humanized form of BTI-322) against CD2 receptor on T-cells, monoclonal antibody LDP-02, muromonab-CD3(OKT3)

antibody, OKT3 ("muromomab-CD3") antibody, PRO 542 antibody, ReoPro ("abciximab") antibody, and TNF-IgG fusion protein.

Antibodies may also be made, in accordance with the present invention, against the following non-limiting list of organisms associated with disease and against peptides and/or proteins or other epitopes from these organisms: Anthrax (*Bacillus anthracis*), *Bacillus anthracis* (anthrax), Botulism (*Clostridium botulinum* toxin), *Brucella* species (brucellosis), Brucellosis (*Brucella* species), *Burkholderia mallei* (glanders), *Burkholderia pseudomallei* (melioidosis), *Chlamydia psittaci* (psittacosis), Cholera (*Vibrio cholerae*), *Clostridium botulinum* toxin (botulism), *Clostridium perfringens* (Epsilon toxin), *Coxiella burnetii* (Q fever), *E. coli* O157:H7 (*Escherichia coli*), Emerging infectious diseases such as Nipah virus and hantavirus, Epsilon toxin of *Clostridium perfringens*, *Escherichia coli* O157:H7 (*E. coli*), Food safety threats (e.g., *Salmonella* species, *Escherichia coli* O157:H7, *Shigella*, *Francisella tularensis* (tularemia), Glanders (*Burkholderia mallei*), Melioidosis (*Burkholderia pseudomallei*), Plague (*Yersinia pestis*), Psittacosis (*Chlamydia psittaci*), Q fever (*Coxiella burnetii*), Ricin toxin from *Ricinus communis* (castor beans), *Rickettsia prowazekii* (typhus fever), *Salmonella* species (salmonellosis), *Salmonella* Typhi (typhoid fever), Salmonellosis (*Salmonella* species), *Shigella* (shigellosis), Shigellosis (*Shigella*), Smallpox (variola major), Staphylococcal enterotoxin B, Tularemia (*Francisella tularensis*), Typhoid fever (*Salmonella* Typhi), Typhus fever (*Rickettsia prowazekii*), Variola major (smallpox), *Vibrio cholerae* (cholera), Viral encephalitis (alphaviruses [e.g., Venezuelan equine encephalitis, eastern equine encephalitis, western equine encephalitis]), Viral hemorrhagic fevers (filoviruses [e.g., Ebola, Marburg] and arenaviruses [e.g., Lassa, Machupo]), Water safety threats (e.g., *Vibrio cholerae*, *Cryptosporidium parvum*), and *Yersinia pestis* (plague).

Another non-limiting list of the antibodies that may be produced using the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. (www.phoenixpeptide.com; 530 Harbor Boulevard, Belmont, CA), Peninsula Labs San Carlos CA, SIGMA, St.Louis, MO www.sigma-aldrich.com, Cappel ICN, Irvine, California, www.icnbiomed.com, and Calbiochem, La Jolla, California, www.calbiochem.com, which are all incorporated herein by reference in their entirety. The polynucleotide sequences encoding these antibodies

may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the antibody polynucleotide sequence by choosing the codons that encode for each amino acid in the desired antibody.

5       As an example of protein production using the method of the present invention which may be applied to the production of proteins useful in vaccines, the gene of interest is a proinsulin gene and the desired molecule is insulin. Proinsulin consists of three parts: a C-peptide and two strands of amino acids (the alpha and beta chains) that later become linked together to form the insulin molecule. In these embodiments,  
10       proinsulin is expressed in the oviduct tubular gland cells and then deposited in the egg white. One example of a proinsulin polynucleotide sequence is shown in SEQ ID NO:80, wherein the C-peptide cleavage site spans from Arg at position 31 to Arg at position 65.

      Further included in the present invention are genes of interest that encode  
15       proteins and peptides synthesized by the immune system including those synthesized by the thymus, lymph nodes, spleen, and the gastrointestinal associated lymph tissues (GALT) system. The immune system proteins and peptides proteins that can be made in transgenic animals using the polynucleotide cassettes of the present invention include, but are not limited to, alpha-interferon, beta-interferon, gamma-interferon,  
20       alpha-interferon A, alpha-interferon 1, G-CSF, GM-CSF, interleukin-1 (IL-1), IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, TNF- $\alpha$ , and TNF- $\beta$ . Other cytokines included in the present invention include cardiotrophin, stromal cell derived factor, macrophage derived chemokine (MDC), melanoma growth stimulatory activity (MGSA), macrophage inflammatory proteins 1 alpha (MIP-1  
25       alpha), 2, 3 alpha, 3 beta, 4 and 5.

      Genes encoding lytic peptides such as p146 are also included in the genes of interest of the present invention. In one embodiment, the p146 peptide comprises an amino acid sequence of SEQ ID NO:78. The present invention also encompasses a polynucleotide cassette comprising a p146 nucleic acid having a sequence of SEQ ID  
30       NO:79.

      Enzymes are another class of proteins that may be encoded by the polynucleotide cassettes of the present invention. Such enzymes include but are not limited to adenosine deaminase, alpha-galactosidase, cellulase, collagenase, dnaseI,

hyaluronidase, lactase, L-asparaginase, pancreatin, papain, streptokinase B, subtilisin, superoxide dismutase, thrombin, trypsin, urokinase, fibrinolysin, glucocerebrosidase and plasminogen activator. In some embodiments wherein the enzyme could have deleterious effects, additional amino acids and a protease cleavage site are added to the carboxy end of the enzyme of interest in order to prevent expression of a functional enzyme. Subsequent digestion of the enzyme with a protease results in activation of the enzyme. Preferred are enzymes associated with disease processes and the function of infectious organisms.

A non-limiting list of the peptides and proteins that may be encoded by the polynucleotide cassettes of the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. ([www.phoenixpeptide.com](http://www.phoenixpeptide.com); 530 Harbor Boulevard, Belmont, CA), Peninsula Labs (San Carlos CA), SIGMA, (St. Louis, MO [www.sigma-aldrich.com](http://www.sigma-aldrich.com)), Cappel ICN (Irvine, California, [www.icnbiomed.com](http://www.icnbiomed.com)), and Calbiochem (La Jolla, California, [www.calbiochem.com](http://www.calbiochem.com)). The polynucleotide sequences encoding these proteins and peptides of interest may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the polynucleotide sequence to be incorporated into the genome by choosing the codons that encode for each amino acid in the desired protein or peptide.

Other desired proteins that may be encoded by the polynucleotide cassettes of the present invention include bacitracin, polymixin b, vancomycin, cyclosporine, anti-RSV antibody, alpha-1 antitrypsin (AAT), anti-cytomegalovirus antibody, anti-hepatitis antibody, anti-inhibitor coagulant complex, anti-rabies antibody, anti-Rh(D) antibody, adenosine deaminase, anti-digoxin antibody, antivenin crotalidae (rattlesnake venom antibody), antivenin latrodectus (black widow spider venom antibody), antivenin micrurus (coral snake venom antibody), aprotinin, corticotropin (ACTH), diphtheria antitoxin, lymphocyte immune globulin (anti-thymocyte antibody), protamine, thyrotropin, capreomycin,  $\alpha$ -galactosidase, gramicidin, streptokinase, tetanus toxoid, tyrothricin, IGF-1, proteins of varicella vaccine, anti-TNF antibody, anti-IL-2r antibody, anti-HER-2 antibody, OKT3 ("muromonab-CD3") antibody, TNF-IgG fusion protein, ReoPro ("abciximab") antibody, ACTH fragment 1-24, desmopressin, gonadotropin-releasing hormone, histrelin, leuprolide, lypressin, nafarelin, peptide that binds GPIIb/GPIIIa on platelets (integrilin),

goserelin, capreomycin, colistin, anti-respiratory syncytial virus, lymphocyte immune globulin (Thymoglovin, Atgam), panorex, alpha-antitrypsin, botulinin, lung surfactant protein, tumor necrosis receptor-IgG fusion protein (enbrel), gonadorelin, proteins of influenza vaccine, proteins of rotavirus vaccine, proteins of haemophilus b conjugate vaccine, proteins of poliovirus vaccine, proteins of pneumococcal conjugate vaccine, proteins of meningococcal C vaccine, proteins of influenza vaccine, megakaryocyte growth and development factor (MGDF), neuroimmunophilin ligand-A (NIL-A), brain-derived neurotrophic factor (BDNF), glial cell line-derived neurotrophic factor (GDNF), leptin (native), leptin B, leptin C, IL-1RA (interleukin-1RA), R-568, novel erythropoiesis-stimulating protein (NESP), humanized mAb to rous sarcoma virus (MEDI-493), glutamyl-tryptophan dipeptide IM862, LFA-3TIP immunosuppressive, humanized anti-CD40-ligand monoclonal antibody (5c8), gelsolin enzyme, tissue factor pathway inhibitor (TFPI), proteins of meningitis B vaccine, antimetastatic cancer antibody (mAb 17-1A), chimeric (human & mouse) mAb against TNF $\alpha$ , mAb against factor VII, relaxin, capreomycin, glycopeptide (LY333328), recombinant human activated protein C (rhAPC), humanized mAb against the epidermal growth receptor-2, altepase, anti-CD20 antigen, C2B8 antibody, insulin-like growth factor-1, atrial natriuretic peptide (anaritide), tenectapase, anti-CD11a antibody (hu 1124), anti-CD18 antibody, mAb LDP-02, anti-VEGF antibody, fab fragment of anti-VEGF Ab, APO2 ligand (tumor necrosis factor-related apoptosis-inducing ligand), rTGF- $\beta$  (transforming growth factor- $\beta$ ), alpha-antitrypsin, ananain (a pineapple enzyme), humanized mAb CTLA4IG, PRO 542 (mAb), D2E7 (mAb), calf intestine alkaline phosphatase,  $\alpha$ -L-iduronidase,  $\alpha$ -L-galactosidase (humanglutamic acid decarboxylase, acid sphingomyelinase, bone morphogenetic protein-2 (rhBMP-2), proteins of HIV vaccine, T cell receptor (TCR) peptide vaccine, TCR peptides, V beta 3 and V beta 13.1. (IR502), (IR501), BI 1050/1272 mAb against very late antigen-4 (VLA-4), C225 humanized mAb to EGF receptor, anti-idiotypic antibody to GD3 glycolipid, antibacterial peptide against *H. pylori*, MDX-447 bispecific humanized mAb to EGF receptor, anti-cytomegalovirus (CMV), Medi-491 B19 parvovirus vaccine, humanized recombinant mAb (IgG1k) against respiratory syncytial virus (RSV), urinary tract infection vaccine (against "pili" on *Escherichia coli* strains), proteins of lyme disease vaccine against *B. burgdorferi* protein (DbpA), proteins of Medi-501 human papilloma virus-11 vaccine (HPV), *Streptococcus pneumoniae* vaccine, Medi-507

mAb (humanized form of BTI-322) against CD2 receptor on T-cells, MDX-33 mAb to Fc $\gamma$ R1 receptor, MDX-RA immunotoxin (ricin A linked) mAb, MDX-210 bi-specific mAb against HER-2, MDX-447 bi-specific mAb against EGF receptor, MDX-22, MDX-220 bi-specific mAb against TAG-72 on tumors, colony-stimulating factor (CSF) (molgramostim), humanized mAb to the IL-2 R  $\alpha$ -chain (basiliximab),  
 5 mAb to IgE (IGE 025A), myelin basic protein-altered peptide (MSP771A), humanized mAb against the epidermal growth receptor-2, humanized mAb against the  $\alpha$  subunit of the interleukin-2 receptor, low molecular weight heparin, anti-hemophillic factor, and bactericidal/permeability-increasing protein (r-BPI).

10 The following paragraph describes specific examples of epitopes and antibodies associated with specific diseases that may be made as the proteins, multimeric proteins of the present invention. Such epitopes may also be used in antibody production. Anthrax: PA (protective antigen) mediates entry of toxin's lethal factor (LF) into cell, therefore PA is binding domain. There are at least three  
 15 different binding sites on PA where LF can bind. Site I epitope IG3PA63, site II 2D3PA, 2D5PA, 10D2PA, and site III 3B6PA, 14B7PA, 10E10PA63. Botulism: BoNT (botulism neurotoxin), binding domains within neurotoxin A (BoNT/A) derived from segments of heavy chain and encoded proteins : H455-661, H1150-1289. Accession number X73844. BoNT/A in infants: Accession number X73423.

20 Tularemia: mAbs against LPS of francisella tularensis: FT14, FT2F11. OMP specific mAbs react to IgG1 subclass of *F. tularensis*. Yersenia pestis: causes pneumonic and bubonic plague. LCrV: accession number M26405 (*Y. pestis*), LCrV: accession number M57893 (*Y. pseudotuberculosis*). Caf-1 encodes capsular antigen fraction 1 (F1) accession number X61996. Shigella: epitopes 2C32E6, 4D64B9  
 25 (reactive to *S. flexneri* and *S. boydii*). 5E45D8 – *S. flexneri*. 4B33D10, 1B52F10 – all species of Shieglla. Shigella monoclonal antibodies 2C32E6, 4D64B9, 5E45D8 (reactive with *S. Felxneri*). 4B33D10 and 1B52F1 (species of Shigella) and 1B52F10 all strains (see Rahman & Stimson, Hybridoma, 20:85-90, 2001). See Venkatesan et al., PNAS 85:9317-21 (1988) accession number J04117 *S. flexnari*). IpaD between  
 30 amino acid residues 14 and 44 of Shigella flexnari. IpaD is against the unique region of this protein exposed on the surface of the pathogen (Turbyfill et al., Infect. Immunity 66:1999-2006 (1998). See Muller-Loennies et al., J. Biol. Chem. 278: 25618-27 (2003) for epitopes related to Shigella. LcrV, Yersinia pestis (accession.



number M26405), *Yersinia pseudotuberculosis* (accessionnumber M57893, *Yersinia enterocolitica*. Fusion of IcrV and the structural gene for protein A. One protective epitope resides internally between amino acids 168 and 275 (Motin et al., Infect. Immunity 62:41920-201 (1994). For additional epitopes on *Yersinia pastis* see  
5 accession number M26405 LcrV (Price et al., J. Bacteriol. 171: 5646-5653 (1989). For additional epitopes on *Yersinia pastis* see cafl encoding cafl capsular antigen fraction 1 (F1) accession number X61996 in Galyov et al., FEBS Lett. 277: 230-2 (1990). 12G12 SLP-M common specific epitope of the *Brucella* S-LPS. S-LPS of *B. melitensis* Rev1, S-LPS smooth lipopolysaccharide, (Weynants et al., Clin. Diagn.  
10 Lab. Immunol. 3:309-14 (1996). Smooth *Brucella* species – two lipopolysaccharide (LPSs (A and M) produce monoclonal antibodies to the specific A- or M-LPS epitopes. 12AE6, and Mab 33.1.5, s-LPSA=12AE6, LPSM=33.1.5, BmE10-5, 12G12 O side chain = BRU-38 (Douglas et al., J. Clin. Microbiol. 26: 1353-6 (1988). BmE10-5 Mantigen of *Brucella melitensis* 16M, *B. melitensis*, *B. abortus*, *B. suis* and  
15 *B. neotomae* (Vizcaino et al., Res. Microbiol. 143: 513-8 (1992). BRU-38, O side chain of *B. abortus* (Roop et al., J. Clin. Microbiol. 25: 2090-3 (1987). The OMP-specific monoclonal antibodies Bhatti et al., Hybridoma 12: 197-202 (1993). Mabs FT14 and FT2F11, *Francisella tularensis* (Fulop et al., J. Clin. Microbiol. 29: 1407-12 (1991).

20 Preferred proteins and multimeric proteins to be made with the present invention include those associated with disease processes and those associated with antigenic epitopes of infectious organisms.

The proteins and multimeric proteins made using the present invention may be labeled using labels and techniques known to one of ordinary skill in the art. Some of  
25 these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety. Some of these labels may be genetically engineered into the polynucleotide sequence for the expression of the selected multimeric protein. The peptides and proteins may also have label-  
30 incorporation "handles" incorporated to allow labeling of an otherwise difficult or impossible to label multimeric protein.

It is to be understood that the various classes of desired peptides and proteins, as well as specific peptides and proteins described in this section may be modified as

described below by inserting selected codons for desired amino acid substitutions into the gene incorporated into the transgenic animal.

The present invention further encompasses the use of inhibitory molecules to inhibit endogenous (i.e., non-vector) protein production. These inhibitory molecules include antisense nucleic acids, siRNA and inhibitory proteins. Preferred inhibitory molecules are those associated with causation or inhibition of disease processes. In a preferred embodiment, the endogenous protein whose expression is inhibited is an egg white protein including, but not limited to ovalbumin, ovotransferrin, and ovomucin ovomucoid, ovoinhibitor, cystatin, ovostatin, lysozyme, ovoglobulin G2, ovoglobulin G3, avidin, and thiamin binding protein. In one embodiment, a polynucleotide cassette containing an ovalbumin DNA sequence, that upon transcription forms a double stranded RNA molecule, is transfected into an animal such as a bird and the bird's production of endogenous ovalbumin protein is reduced by the interference RNA mechanism (RNAi). In other embodiments, a polynucleotide cassette encodes an inhibitory RNA molecule that inhibits the expression of more than one egg white protein. Additionally, inducible knockouts or knockdowns of the endogenous protein may be created to achieve a reduction or inhibition of endogenous protein production. Endogenous egg white production can be inhibited in an avian at any time, but is preferably inhibited preceding, or immediately preceding, the harvest of eggs.

#### 20 Modified Desired Proteins and Peptides

The present invention may be used for the production of proteins and multimeric proteins. "Proteins", "peptides," "polypeptides" and "oligopeptides" are chains of amino acids (typically L-amino acids) whose alpha carbons are linked through peptide bonds formed by a condensation reaction between the carboxyl group of the alpha carbon of one amino acid and the amino group of the alpha carbon of another amino acid. The terminal amino acid at one end of the chain (i.e., the amino terminal) has a free amino group, while the terminal amino acid at the other end of the chain (i.e., the carboxy terminal) has a free carboxyl group. As such, the term "amino terminus" (abbreviated N-terminus) refers to the free alpha-amino group on the amino acid at the amino terminal of the protein, or to the alpha-amino group (imino group when participating in a peptide bond) of an amino acid at any other location within the protein. Similarly, the term "carboxy terminus" (abbreviated C-terminus) refers to

the free carboxyl group on the amino acid at the carboxy terminus of a protein, or to the carboxyl group of an amino acid at any other location within the protein.

Typically, the amino acids making up a protein are numbered in order, starting at the amino terminal and increasing in the direction toward the carboxy terminal of the protein. Thus, when one amino acid is said to "follow" another, that amino acid is positioned closer to the carboxy terminal of the protein than the preceding amino acid.

The term "residue" is used herein to refer to an amino acid (D or L) or an amino acid mimetic that is incorporated into a protein by an amide bond. As such, the amino acid may be a naturally occurring amino acid or, unless otherwise limited, may encompass known analogs of natural amino acids that function in a manner similar to the naturally occurring amino acids (i.e., amino acid mimetics). Moreover, an amide bond mimetic includes peptide backbone modifications well known to those skilled in the art.

Furthermore, one of skill will recognize that, as mentioned above, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than about 5%, more typically less than about 1%) in an encoded sequence are conservatively modified variations where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

A conservative substitution is a substitution in which the substituting amino acid (naturally occurring or modified) is structurally related to the amino acid being substituted, i.e., has about the same size and electronic properties as the amino acid being substituted. Thus, the substituting amino acid would have the same or a similar functional group in the side chain as the original amino acid. A "conservative substitution" also refers to utilizing a substituting amino acid which is identical to the

amino acid being substituted except that a functional group in the side chain is protected with a suitable protecting group.

Suitable protecting groups are described in Green and Wuts, "Protecting Groups in Organic Synthesis", John Wiley and Sons, Chapters 5 and 7, 1991, the teachings of which are incorporated herein by reference. Preferred protecting groups are those which facilitate transport of the peptide through membranes, for example, by reducing the hydrophilicity and increasing the lipophilicity of the peptide, and which can be cleaved, either by hydrolysis or enzymatically (Ditter et al., 1968. J. Pharm. Sci. 57:783; Ditter et al., 1968. J. Pharm. Sci. 57:828; Ditter et al., 1969. J. Pharm. Sci. 58:557; King et al., 1987. Biochemistry 26:2294; Lindberg et al., 1989. Drug Metabolism and Disposition 17:311; Tunek et al., 1988. Biochem. Pharm. 37:3867; Anderson et al., 1985 Arch. Biochem. Biophys. 239:538; and Singhal et al., 1987. FASEB J. 1:220). Suitable hydroxyl protecting groups include ester, carbonate and carbamate protecting groups. Suitable amine protecting groups include acyl groups and alkoxy or aryloxy carbonyl groups, as described above for N-terminal protecting groups. Suitable carboxylic acid protecting groups include aliphatic, benzyl and aryl esters, as described below for C-terminal protecting groups. In one embodiment, the carboxylic acid group in the side chain of one or more glutamic acid or aspartic acid residues in a peptide of the present invention is protected, preferably as a methyl, ethyl, benzyl or substituted benzyl ester, more preferably as a benzyl ester.

Provided below are groups of naturally occurring and modified amino acids in which each amino acid in a group has similar electronic and steric properties. Thus, a conservative substitution can be made by substituting an amino acid with another amino acid from the same group. It is to be understood that these groups are non-limiting, i.e. that there are additional modified amino acids which could be included in each group.

Group I includes leucine, isoleucine, valine, methionine and modified amino acids having the following side chains: ethyl, n-propyl n-butyl. Preferably, Group I includes leucine, isoleucine, valine and methionine.

Group II includes glycine, alanine, valine and a modified amino acid having an ethyl side chain. Preferably, Group II includes glycine and alanine.

Group III includes phenylalanine, phenylglycine, tyrosine, tryptophan, cyclohexylmethyl glycine, and modified amino residues having substituted

benzyl or phenyl side chains. Preferred substituents include one or more of the following: halogen, methyl, ethyl, nitro,  $\text{—NH}_2$ , methoxy, ethoxy and  $\text{—CN}$ . Preferably, Group III includes phenylalanine, tyrosine and tryptophan.

5 Group IV includes glutamic acid, aspartic acid, a substituted or unsubstituted aliphatic, aromatic or benzylic ester of glutamic or aspartic acid (e.g., methyl, ethyl, n-propyl iso-propyl, cyclohexyl, benzyl or substituted benzyl), glutamine, asparagine,  $\text{—CO—NH—}$  alkylated glutamine or asparagines (e.g., methyl, ethyl, n-propyl and iso-propyl) and modified amino acids having the side chain  $\text{—(CH}_2)_3\text{—COOH}$ , an ester thereof (substituted or unsubstituted  
10 aliphatic, aromatic or benzylic ester), an amide thereof and a substituted or unsubstituted N-alkylated amide thereof. Preferably, Group IV includes glutamic acid, aspartic acid, methyl aspartate, ethyl aspartate, benzyl aspartate and methyl glutamate, ethyl glutamate and benzyl glutamate, glutamine and asparagine.

15 Group V includes histidine, lysine, ornithine, arginine, N-nitroarginine,  $\beta$ -cycloarginine,  $\gamma$ -hydroxyarginine, N-amidinocitruline and 2-amino-4-guanidinobutanoic acid, homologs of lysine, homologs of arginine and homologs of ornithine. Preferably, Group V includes histidine, lysine, arginine and ornithine. A homolog of an amino acid includes from 1 to about  
20 3 additional or subtracted methylene units in the side chain.

Group VI includes serine, threonine, cysteine and modified amino acids having C1-C5 straight or branched alkyl side chains substituted with  $\text{—OH}$  or  $\text{—SH}$ , for example,  $\text{—CH}_2\text{CH}_2\text{OH}$ ,  $\text{—CH}_2\text{CH}_2\text{CH}_2\text{OH}$  or  $\text{—CH}_2\text{CH}_2\text{OHCH}_3$ . Preferably, Group VI includes serine, cysteine or threonine.

25 In another aspect, suitable substitutions for amino acid residues include “severe” substitutions. A “severe substitution” is a substitution in which the substituting amino acid (naturally occurring or modified) has significantly different size and/or electronic properties compared with the amino acid being substituted. Thus, the side chain of the substituting amino acid can be significantly larger (or  
30 smaller) than the side chain of the amino acid being substituted and/or can have functional groups with significantly different electronic properties than the amino acid being substituted. Examples of severe substitutions of this type include the substitution of phenylalanine or cyclohexylmethyl glycine for alanine, isoleucine for

glycine, a D amino acid for the corresponding L amino acid, or  $\text{—NH—CH[(—CH}_2\text{)}_5\text{—COOH]—CO—}$  for aspartic acid. Alternatively, a functional group may be added to the side chain, deleted from the side chain or exchanged with another functional group. Examples of severe substitutions of this type include adding of  
5 valine, leucine or isoleucine, exchanging the carboxylic acid in the side chain of aspartic acid or glutamic acid with an amine, or deleting the amine group in the side chain of lysine or ornithine. In yet another alternative, the side chain of the substituting amino acid can have significantly different steric and electronic properties that the functional group of the amino acid being substituted. Examples of such  
10 modifications include tryptophan for glycine, lysine for aspartic acid and  $\text{—(CH}_2\text{)}_4\text{COOH}$  for the side chain of serine. These examples are not meant to be limiting.

In another embodiment, for example in the synthesis of a peptide 26 amino acids in length, the individual amino acids may be substituted according in the  
15 following manner:

- AA<sub>1</sub> is serine, glycine, alanine, cysteine or threonine;  
AA<sub>2</sub> is alanine, threonine, glycine, cysteine or serine;  
AA<sub>3</sub> is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine,  $\beta$ -cycloarginine,  $\gamma$ -hydroxyarginine, N-amidinocitruline or 2-amino-4-  
20 guanidinobutanoic acid;  
AA<sub>4</sub> is proline, leucine, valine, isoleucine or methionine;  
AA<sub>5</sub> is tryptophan, alanine, phenylalanine, tyrosine or glycine;  
AA<sub>6</sub> is serine, glycine, alanine, cysteine or threonine;  
AA<sub>7</sub> is proline, leucine, valine, isoleucine or methionine;  
25 AA<sub>8</sub> is alanine, threonine, glycine, cysteine or serine;  
AA<sub>9</sub> is alanine, threonine, glycine, cysteine or serine;  
AA<sub>10</sub> is leucine, isoleucine, methionine or valine;  
AA<sub>11</sub> is serine, glycine, alanine, cysteine or threonine;  
AA<sub>12</sub> is leucine, isoleucine, methionine or valine;  
30 AA<sub>13</sub> is leucine, isoleucine, methionine or valine;  
AA<sub>14</sub> is glutamine, glutamic acid, aspartic acid, asparagine, or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

- AA<sub>15</sub> is arginine, N-nitroarginine,  $\beta$ -cycloarginine,  $\gamma$ -hydroxy-arginine, N-amidinocitruline or 2-amino-4-guanidino-butanoic acid
- AA<sub>16</sub> is proline, leucine, valine, isoleucine or methionine;
- AA<sub>17</sub> is serine, glycine, alanine, cysteine or threonine;
- 5 AA<sub>18</sub> is glutamic acid, aspartic acid, asparagine, glutamine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
- AA<sub>19</sub> is aspartic acid, asparagine, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
- 10 AA<sub>20</sub> is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine,  $\beta$ -cycloarginine,  $\gamma$ -hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;
- AA<sub>21</sub> is alanine, threonine, glycine, cysteine or serine;
- AA<sub>22</sub> is alanine, threonine, glycine, cysteine or serine;
- 15 AA<sub>23</sub> is histidine, serine, threonine, cysteine, lysine or ornithine;
- AA<sub>24</sub> is threonine, aspartic acid, serine, glutamic acid or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
- AA<sub>25</sub> is asparagine, aspartic acid, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of
- 20 glutamic acid or aspartic acid; and
- AA<sub>26</sub> is cysteine, histidine, serine, threonine, lysine or ornithine.

It is to be understood that these amino acid substitutions may be made for longer or shorter peptides than the 26 mer in the preceding example above, and for proteins.

- 25 In one embodiment of the present invention, codons for the first several N-terminal amino acids of the transposase are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. It is preferable that between approximately 1 and 20, more preferably 3 and 15, and most preferably between 4 and 12 of the first N-terminal codons of the gene of interest are
- 30 modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. In one embodiment, the first ten N-terminal codons of the gene of interest are modified in this manner.

When several desired proteins, protein fragments or peptides are encoded in the gene of interest to be incorporated into the genome, as with the multivalent multimeric proteins, one of skill in the art will appreciate that the proteins, protein fragments or peptides may be separated by a spacer molecule such as, for example, a peptide, consisting of one or more amino acids. Generally, the spacer will have no specific biological activity other than to join the desired proteins, protein fragments or peptides together, or to preserve some minimum distance or other spatial relationship between them. However, the constituent amino acids of the spacer may be selected to influence some property of the molecule such as the folding, net charge, or hydrophobicity. The spacer may also be contained within a nucleotide sequence with a purification handle or be flanked by proteolytic cleavage sites.

Such polypeptide spacers may have from about 1 to about 100 amino acids, preferably 3 to 20 amino acids, and more preferably 4-15 amino acids. The spacers in a polypeptide are independently chosen, but are preferably all the same. The spacers should allow for flexibility of movement in space and are therefore typically rich in small amino acids, for example, glycine, serine, proline or alanine. Preferably, peptide spacers contain at least 60%, more preferably at least 80% glycine or alanine. In addition, peptide spacers generally have little or no biological and antigenic activity. Preferred spacers are (Gly-Pro-Gly-Gly)<sub>x</sub> (SEQ ID NO:81) and (Gly<sub>4</sub>-Ser)<sub>y</sub>, wherein x is an integer from about 3 to about 9 and y is an integer from about 1 to about 8. Specific examples of suitable spacers include

(Gly-Pro-Gly-Gly)<sub>3</sub>

SEQ ID NO:82 Gly Pro Gly Gly Gly Pro Gly Gly Gly Pro Gly Gly

(Gly<sub>4</sub>-Ser)<sub>3</sub>

SEQ ID NO:83 Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser  
or (Gly<sub>4</sub>-Ser)<sub>4</sub>

SEQ ID NO:84 Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser  
Gly Gly Gly Gly Ser.

One example of a multivalent multimeric protein containin a spacer is leutinizing hormone (LH), normally made as separate alpha and beta chains, made as a single polypeptide as described in Galet et. al., Mol. Cell Endocrinology, 2001, 174 (1-2):31-40. Production of a multimeric protein may thus be simplified using a spacer sequence that may or may not contain cleavage sites. In the case of an



immunoglobulin, for example, a heavy and light chain may be synthesized as a single polypeptide using a spacer sequence with protease sites native to the transgenic animal so as to make, upon processing, a heavy and light chain combination in close association, facilitating the addition of a similar heavy and light chain to produce the native immunoglobulin. In this model, the removal of the spacer sequence may or may not be required. Other multimeric proteins may be made in bioengineered organisms in a similar fashion.

Nucleotide sequences encoding for the production of residues which may be useful in purification of the expressed recombinant protein may also be built into the vector. Such sequences are known in the art and include the glutathione binding domain from glutathione S-transferase, polylysine, hexa-histidine or other cationic amino acids, thioredoxin, hemagglutinin antigen and maltose binding protein.

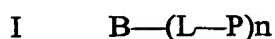
Additionally, nucleotide sequences may be inserted into the gene of interest to be incorporated so that the protein or peptide can also include from one to about six amino acids that create signals for proteolytic cleavage. In this manner, if a gene is designed to make one or more peptides or proteins of interest in the transgenic animal, specific nucleotide sequences encoding for amino acids recognized by enzymes may be incorporated into the gene to facilitate cleavage of the large protein or peptide sequence into desired peptides or proteins or both. For example, nucleotides encoding a proteolytic cleavage site can be introduced into the gene of interest so that a signal sequence can be cleaved from a protein or peptide encoded by the gene of interest. Nucleotide sequences encoding other amino acid sequences which display pH sensitivity, chemical sensitivity or photolability may also be added to the vector to facilitate separation of the signal sequence from the peptide or protein of interest.

Proteolytic cleavage sites include cleavage sites recognized by exopeptidases such as carboxypeptidase A, carboxypeptidase B, aminopeptidase I, and dipeptidylaminopeptidase; endopeptidases such as trypsin, V8-protease, enterokinase, factor Xa, collagenase, endoproteinase, subtilisin, and thrombin; and proteases such as Protease 3C IgA protease (Igase) Rhinovirus 3C(preScission)protease. Chemical cleavage sites are also included in the definition of cleavage site as used herein. Chemical cleavage sites include, but are not limited to, site cleaved by cyanogen bromide, hydroxylamine, formic acid, and acetic acid. Self-splicing cleavage sites such as inteins are also included in the present invention.

In some embodiments, one or more cleavage sites are incorporated into a polynucleotide cassette containing multiple genes of interest. Figure 4 depicts one example of a polynucleotide cassette containing two genes of interest containing a cleavage site between them. The genes of interest may encode different proteins or peptides, the same protein or peptide, or modified versions of the same protein or peptide. While Figure 4 shows a polynucleotide cassette containing two genes of interest, the present invention encompasses a polynucleotide cassette containing any number of genes of interest. The cleavage site located between the genes of interest can encode any amino acid sequence that is cleaved by any means. As mentioned above, the cleavage site can encode an amino acid sequence cleaved by a protease, a chemical reaction, can be a photolabile site, or can be a pro polynucleotide.

The present invention includes a polynucleotide cassette that encodes a repetitive polypeptide chain in which two or more peptides, polypeptides or proteins, designated as P in the structural formulae presented below, are each separated by a peptide spacer or cleavage site designated as B. A polypeptide multivalent ligand, also called a multivalent protein, is a form of a multimeric protein encoded by the polynucleotide cassettes of the present invention, and is represented by structural formulae (I, II and III). Each peptide or protein is connected to another peptide or protein through a peptide bond, to a linker group, to a spacer, or to a cleavage site. Each peptide, polypeptide or protein may be the same or different and each linker, spacer, cleavage site or covalent bond is independently chosen.

A "polypeptide multivalent protein" is a multiple repeat polypeptide chain in which two or more peptides P are each separated by a peptide linker group, a spacer or a cleavage site. A polypeptide multivalent ligand is represented by structural formulae II and III.



wherein B is a peptide spacer or cleavage site, n is an integer from 2 to about 20, each L is a covalent bond, a linking group or cleavage site which may be present or absent, and each P is a peptide having from about 4 to about 200 amino acid residues.

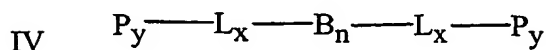


wherein m is an integer from 0 to about 20.

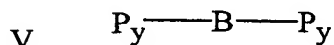
### III Pa - (B)<sub>n</sub> - Pa

- 5 wherein n is an integer from 1 to 20, preferably 2 to 10, more preferably 3 to 7, further wherein a is 1.

Other examples of multivalent proteins include the following:



10



In the preceding structural formulae IV and V of polypeptide multivalent ligands encoded by a polynucleotide cassette of the present invention, each P is a peptide having from about 4 to about 200 amino acid residues, y is 1, x is an integer from 1 to 3, and n is an integer from 1 to 20, preferably 2 to 10, more preferably 3 to 7. Each B is a peptide spacer or cleavage site comprised of at least 2 amino acids or a cleavage site. Each peptide P and each B are independently chosen and may be the same or different.

Suitable linkers (L) are groups that can connect peptides and proteins to each other. In one example, the linker is an oligopeptide of from about 1 to about 10 amino acids consisting of amino acids with inert side chains. Suitable oligopeptides include polyglycine, polyserine, polyproline, polyalanine and oligopeptides consisting of alanyl and/or serinyl and/or prolinyl and/or glycyl amino acid residues. m in structural formula II is an integer from 0 to about 20.

25 The peptides, polypeptides and proteins in a multivalent protein can be connected to each other by covalent bonds, linker groups, spacers, cleavage groups or a combination thereof. The linking groups can be the same or different.

A polypeptide spacer shown in structural formula (II) is a peptide having from about 5 to about 40 amino acid residues. The spacers in a polypeptide multivalent ligand are independently chosen, and may be the same or different. The spacers should allow for flexibility of movement in space for the flanking peptides, polypeptides and proteins P, and are therefore typically rich in small amino acids, for

example, glycine, serine, proline or alanine. Preferably, peptide spacers contain at least 60%, more preferably at least 80% glycine or alanine. In addition, peptide spacers generally have little or no biological and antigenic activity. Preferred spacers are (Gly-Pro-Gly-Gly)<sub>x</sub> (SEQ ID NO:81) and (Gly<sub>4</sub>-Ser)<sub>y</sub>, wherein x is an integer  
5 from about 3 to about 9 and y is an integer from about 1 to about 8. Specific examples of suitable spacers include (Gly<sub>4</sub>-Ser)<sub>3</sub> (SEQ ID NO:82). Spacers can also include from one to about four amino acids that create a signal for proteolytic cleavage.

In another embodiment of the present invention, a TAG sequence is linked to a  
10 gene of interest. The TAG sequence serves three purposes: 1) it allows free rotation of the peptide or protein to be isolated so there is no interference from the native protein or signal sequence, i.e. vitellogenin, 2) it provides a "purification handle" to isolate the protein using affinity purification, and 3) it includes a cleavage site to remove the desired protein from the signal and purification sequences. Accordingly,  
15 as used herein, a TAG sequence includes a spacer sequence, a purification handle and a cleavage site. The spacer sequences in the TAG proteins contain one or more repeats shown in SEQ ID NO:85. A preferred spacer sequence comprises the sequence provided in SEQ ID NO:86. One example of a purification handle is the gp41 hairpin loop from HIV I. Exemplary gp41 polynucleotide and polypeptide  
20 sequences are provided in SEQ ID NO:87 and SEQ ID NO:88, respectively. However, it should be understood that any antigenic region, or otherwise associative regions such as avidin/biotin, may be used as a purification handle, including any antigenic region of gp41. Preferred purification handles are those that elicit highly specific antibodies. Additionally, the cleavage site can be any protein cleavage site  
25 known to one of ordinary skill in the art and includes an enterokinase cleavage site comprising the Asp Asp Asp Asp Lys sequence (SEQ ID NO:89) and a furin cleavage site. In one embodiment of the present invention, the TAG sequence comprises a polynucleotide sequence of SEQ ID NO:90.

#### Methods of Administering Polynucleotide Cassettes

30 In addition to the polynucleotide cassettes described above, the present invention also includes methods of administering the polynucleotide cassettes to an animal, methods of producing a transgenic animal wherein a gene of interest is incorporated into the germline of the animal and methods of producing a transgenic

animal wherein a gene of interest is incorporated into cells other than the germline cells of the animal. The polynucleotide cassettes may reside in any vector or delivery solution when administered or may be naked DNA. In one embodiment, a transposon-based vector containing the polynucleotide cassette between two insertion  
5 sequences recognized by a transposase is administered to an animal. The polynucleotide cassettes of the present invention may be administered to an animal via any method known to those of skill in the art, including, but not limited to, intraembryonic, intratesticular, intraoviduct, intraovarian, into the duct system of the mammary gland, intraperitoneal, intraarterial, intravenous, topical, oral, nasal, and  
10 pronuclear injection methods of administration, or any combination thereof. The polynucleotide cassettes may also be administered within the lumen of an organ, into an organ, into a body cavity, into the cerebrospinal fluid, through the urinary system, through the genitourinary system, through the reproductive system, or through any route to reach the desired cells.

15 The polynucleotide cassettes may be delivered through the vascular system to be distributed to the cells supplied by that vessel. For example, the compositions may be placed in the artery supplying the ovary or supplying the fallopian tube to transfect cells in those tissues. In this manner, follicles could be transfected to create a germline transgenic animal. Alternatively, supplying the compositions through the  
20 artery leading to the oviduct would preferably transfect the tubular gland and epithelial cells. Such transfected cells could manufacture a desired protein or peptide for deposition in the egg white. Administration of the compositions through the portal vein would target uptake and transformation of hepatic cells. Administration through the urethra and into the bladder would target the transitional epithelium of the bladder.  
25 Administration through the vagina and cervix would target the lining of the uterus and the epithelial cells of the fallopian tube. Administration through the internal mammary artery or through the duct system of the mammary gland would transfect secretory cells of the lactating mammary gland to perform a desired function, such as to synthesize and secrete a desired protein or peptide into the milk.

30 The polynucleotide cassettes may be administered in a single administration, multiple administrations, continuously, or intermittently. The polynucleotide cassettes may be administered by injection, via a catheter, an osmotic mini-pump or any other method. In some embodiments, a polynucleotide cassette is administered to

an animal in multiple administrations, each administration containing the polynucleotide cassette and a different transfecting reagent.

In a preferred embodiment, the animal is an egg-laying animal, and more preferably, an avian. In one embodiment, between approximately 1 and 150  $\mu\text{g}$ , 1 and 100  $\mu\text{g}$ , 1 and 50  $\mu\text{g}$ , preferably between 1 and 20  $\mu\text{g}$ , and more preferably between 5 and 10  $\mu\text{g}$  of a transposon-based vector containing the polynucleotide cassette is administered to the oviduct of a bird. In a chicken, it is preferred that between approximately 1 and 100  $\mu\text{g}$ , or 5 and 50  $\mu\text{g}$  are administered. In a quail, it is preferred that between approximately 5 and 10  $\mu\text{g}$  are administered. Optimal ranges depending upon the type of bird and the bird's stage of sexual maturity. Intraoviduct administration of the transposon-based vectors of the present invention result in a PCR positive signal in the oviduct tissue, whereas intravascular administration results in a PCR positive signal in the liver. In other embodiments, the polynucleotide cassettes is administered to an artery that supplies the oviduct or the liver. These methods of administration may also be combined with any methods for facilitating transfection, including without limitation, electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

The transposon-based vectors may be administered to the animal at any point during the lifetime of the animal, however, it is preferable that the vectors are administered prior to the animal reaching sexual maturity. The transposon-based vectors are preferably administered to a chicken oviduct between approximately 14 and 16 weeks of age and to a quail oviduct between approximately 5 and 10 weeks of age, more preferably 5 and 8 weeks of age, and most preferably between 5 and 6 weeks of age, when standard poultry rearing practices are used. The vectors may be administered at earlier ages when exogenous hormones are used to induce early sexual maturation in the bird. In some embodiments, the transposon-based vector is administered to an animal's oviduct following an increase in proliferation of the oviduct epithelial cells and/or the tubular gland cells. Such an increase in proliferation normally follows an influx of reproductive hormones in the area of the oviduct. When the animal is an avian, the transposon-based vector is administered to the avian's oviduct following an increase in proliferation of the oviduct epithelial cells and before the avian begins to produce egg white constituents.

The present invention also includes a method of intraembryonic administration of a transposon-based vector containing a polynucleotide cassette to an avian embryo comprising the following steps: 1) incubating an egg on its side at room temperature for two hours to allow the embryo contained therein to move to top dead center (TDC); 2) drilling a hole through the shell without penetrating the underlying shell membrane; 3) injecting the embryo with the transposon-based vector in solution; 4) sealing the hole in the egg; and 5) placing the egg in an incubator for hatching. Administration of the transposon-based vector can occur anytime between immediately after egg lay (when the embryo is at Stage X) and hatching. Preferably, the transposon-based vector is administered between 1 and 7 days after egg lay, more preferably between 1 and 2 days after egg lay. The transposon-based vectors may be introduced into the embryo in amounts ranging from about 5.0  $\mu\text{g}$  to 10 pg, preferably 1.0  $\mu\text{g}$  to 100 pg. Additionally, the transposon-based vector solution volume may be between approximately 1  $\mu\text{l}$  to 75  $\mu\text{l}$  in quail and between approximately 1  $\mu\text{l}$  to 500  $\mu\text{l}$  in chicken.

The present invention also includes a method of intratesticular administration of a transposon-based vector containing a polynucleotide cassette including injecting a bird with a composition comprising the transposon-based vector, an appropriate carrier and an appropriate transfection reagent. In one embodiment, the bird is injected before sexual maturity, preferably between approximately 4-14 weeks, more preferably between approximately 6-14 weeks and most preferably between 8-12 weeks old. In another embodiment, a mature bird is injected with a transposon-based vector an appropriate carrier and an appropriate transfection reagent. The mature bird may be any type of bird, but in one example the mature bird is a quail.

A bird is preferably injected prior to the development of the blood-testis barrier, which thereby facilitates entry of the transposon-based vector into the seminiferous tubules and transfection of the spermatogonia or other germline cells. At and between the ages of 4, 6, 8, 10, 12, and 14 weeks, it is believed that the testes of chickens are likely to be most receptive to transfection. In this age range, the blood/testis barrier has not yet formed, and there is a relatively high number of spermatogonia relative to the numbers of other cell types, e.g., spermatids, etc. See J. Kumaran et al., 1949. Poultry Sci., 29:511-520. See also E. Oakberg, 1956. Am. J. Anatomy, 99:507-515; and P. Kluin et al., 1984. Anat. Embryol., 169:73-78.

The transposon-based vectors may be introduced into a testis in an amount ranging from about 0.1  $\mu$ g to 10  $\mu$ g, preferably 1  $\mu$ g to 10  $\mu$ g, more preferably 3  $\mu$ g to 10  $\mu$ g. In a quail, about 5  $\mu$ g is a preferred amount. In a chicken, about 5  $\mu$ g to 10  $\mu$ g per testis is preferred. These amounts of vector DNA may be injected in one dose or  
5 multiple doses and at one site or multiple sites in the testis. In a preferred embodiment, the vector DNA is administered at multiple sites in a single testis, both testes being injected in this manner. In one embodiment, injection is spread over three injection sites: one at each end of the testis, and one in the middle. Additionally, the transposon-based vector solution volume may be between approximately 1  $\mu$ l to  
10 75  $\mu$ l in quail and between approximately 1  $\mu$ l to 500  $\mu$ l in chicken. In a preferred embodiment, the transposon-based vector solution volume may be between approximately 20  $\mu$ l to 60  $\mu$ l in quail and between approximately 50  $\mu$ l to 250  $\mu$ l in chicken. Both the amount of vector DNA and the total volume injected into each testis may be determined based upon the age and size of the bird.

15 According to the present invention, the polynucleotide cassette is administered in conjunction with an acceptable carrier and/or transfection reagent. Acceptable carriers include, but are not limited to, water, saline, Hanks Balanced Salt Solution (HBSS), Tris-EDTA (TE) and lyotropic liquid crystals. Transfection reagents commonly known to one of ordinary skill in the art that may be employed include, but  
20 are not limited to, the following: cationic lipid transfection reagents, cationic lipid mixtures, polyamine reagents, liposomes and combinations thereof; SUPERFECT®, Cytofectene, BioPORTER®, GenePORTER®, NeuroPORTER®, and perfectin from Gene Therapy Systems; lipofectamine, cellfectin, DMRIE-C oligofectamine, and PLUS reagent from InVitrogen; Xtreme gene, fugene, DOSPER and DOTAP from  
25 Roche; Lipotaxi and Genejammer from Strategene; and Escort from SIGMA. In one embodiment, the transfection reagent is SUPERFECT®. The ratio of DNA to transfection reagent may vary based upon the method of administration. In one embodiment, a transposon-based vector containing a polynucleotide cassette is administered intratesticularly and the ratio of DNA to transfection reagent can be  
30 from 1:1.5 to 1:15, preferably 1:2 to 1:10, all expressed as wt/vol. Transfection may also be accomplished using other means known to one of ordinary skill in the art, including without limitation electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).



Depending upon the cell or tissue type targeted for transfection, the form of the transposon-based vector may be important. Plasmids harvested from bacteria are generally closed circular supercoiled molecules, and this is the preferred state of a vector for gene delivery because of the ease of preparation. In some instances, transposase expression and insertion may be more efficient in a relaxed, closed circular configuration or in a linear configuration. In still other instances, a purified transposase protein may be co-injected with a transposon-based vector containing the gene of interest for more immediate insertion. This could be accomplished by using a transfection reagent complexed with both the purified transposase protein and the transposon-based vector.

#### Testing for and Breeding Animals Carrying the Transgene

Following administration of a polynucleotide cassette to an animal, DNA is extracted from the animal to confirm integration of the genes of interest. Advantages provided by the present invention include the high rates of integration, or incorporation, and transcription of the gene of interest when administered to a bird via an intraoviduct or intraovary route (including intraarterial administrations to arteries leading to the oviduct or ovary) and contained within a transposon-based vector.

Actual frequencies of integration can be estimated both by comparative strength of the PCR signal, and by histological evaluation of the tissues by quantitative PCR. Another method for estimating the rate of transgene insertion is the so-called primed in situ hybridization technique (PRINS). This method determines not only which cells carry a transgene of interest, but also into which chromosome the gene has inserted, and even what portion of the chromosome. Briefly, labeled primers are annealed to chromosome spreads (affixed to glass slides) through one round of PCR, and the slides are then developed through normal in situ hybridization procedures. This technique combines the best features of in situ PCR and fluorescence in situ hybridization (FISH) to provide distinct chromosome location and copy number of the gene in question. The 28s rRNA gene will be used as a positive control for spermatogonia to confirm that the technique is functioning properly. Using different fluorescent labels for the transgene and the 28s gene causes cells containing a transgene to fluoresce with two different colored tags.

Breeding experiments may also be conducted to determine if germline transmission of the transgene has occurred. In a general bird breeding experiment

performed according to the present invention, each male bird is exposed to 2-3 different adult female birds for 3-4 days each. This procedure is continued with different females for a total period of 6-12 weeks. Eggs are collected daily for up to 14 days after the last exposure to the transgenic male, and each egg is incubated in a standard incubator. The resulting embryos are examined for transgene presence at day 3 or 4 using PCR.

Any male producing a transgenic embryo is bred to additional females. Eggs from these females are incubated, hatched, and the chicks tested for the exogenous DNA. Any embryos that die are necropsied and examined directly for the transgene or protein encoded by the transgene, either by fluorescence or PCR. The offspring that hatch and are found to be positive for the exogenous DNA are raised to maturity. These birds are bred to produce further generations of transgenic birds, to verify efficiency of the transgenic procedure and the stable incorporation of the transgene into the germ line. The resulting embryos are examined for transgene presence at day 3 or 4 using PCR.

It is to be understood that the above procedure can be modified to suit animals other than birds and that selective breeding techniques may be performed to amplify gene copy numbers and protein output.

#### Production of Desired Proteins and Multimeric Proteins in Egg White

In one embodiment, a transposon-based vector containing a polynucleotide cassette of the present invention may be administered to a bird for production of desired proteins or peptides in the egg white. These transposon-based vectors preferably contain one or more of an ovalbumin promoter, an ovomucoid promoter, an ovalbumin signal sequence and an ovomucoid signal sequence. Oviduct-specific ovalbumin promoters are described in B. O'Malley et al., 1987. EMBO J., vol. 6, pp. 2305-12; A. Qiu et al., 1994. Proc. Nat. Acad. Sci. (USA), vol. 91, pp. 4451-4455; D. Monroe et al., 2000. Biochim. Biophys. Acta, 1517 (1):27-32; H. Park et al., 2000. Biochem., 39:8537-8545; and T. Muramatsu et al., 1996. Poult. Avian Biol. Rev., 6:107-123.

#### Production of Desired Proteins and Multimeric Proteins in Egg Yolk

The present invention is particularly advantageous for production of recombinant peptides and proteins of low solubility in the egg yolk. Such proteins include, but are not limited to, membrane-associated or membrane-bound proteins,

lipophilic compounds; attachment factors, receptors, and components of second messenger transduction machinery. Low solubility peptides and proteins are particularly challenging to produce using conventional recombinant protein production techniques (cell and tissue cultures) because they aggregate in water-based, hydrophilic environments. Such aggregation necessitates denaturation and re-folding of the recombinantly-produced proteins, which may deleteriously affect their structure and function. Moreover, even highly soluble recombinant peptides and proteins may precipitate and require denaturation and renaturation when produced in sufficiently high amounts in recombinant protein production systems. The present invention provides an advantageous resolution of the problem of protein and peptide solubility during production of large amounts of recombinant proteins.

In one embodiment of the present invention, deposition of a desired protein into the egg yolk is accomplished by attaching a sequence encoding a protein capable of binding to the yolk vitellogenin receptor to a gene of interest that encodes a desired protein. This polynucleotide cassette can be used for the receptor-mediated uptake of the desired protein by the oocytes. In a preferred embodiment, the sequence ensuring the binding to the vitellogenin receptor is a targeting sequence of a vitellogenin protein. The invention encompasses various vitellogenin proteins and their targeting sequences. In a preferred embodiment, a chicken vitellogenin protein targeting sequence is used, however, due to the high degree of conservation among vitellogenin protein sequences and known cross-species reactivity of vitellogenin targeting sequences with their egg-yolk receptors, other vitellogenin targeting sequences can be substituted. One example of a construct for use in the transposon-based vectors of the present invention and for deposition of an insulin protein in an egg yolk is a transposon-based vector containing a vitellogenin promoter, a vitellogenin targeting sequence, a TAG sequence, a pro-insulin sequence and a synthetic polyA sequence. The present invention includes, but is not limited to, vitellogenin targeting sequences residing in the N-terminal domain of vitellogenin, particularly in lipovitellin I. In one embodiment, the vitellogenin targeting sequence contains the polynucleotide sequence of SEQ ID NO:77.

In a preferred embodiment, the transposon-based vector contains a transposase gene operably-linked to a constitutive promoter and a gene of interest operably-linked to a liver-specific promoter and a vitellogenin targeting sequence.

### Isolation and Purification of Desired Proteins and Multimeric Proteins

For large-scale production of protein, an animal breeding stock that is homozygous for the transgene is preferred. Such homozygous individuals are obtained and identified through, for example, standard animal breeding procedures or  
5 PCR protocols.

Once expressed, peptides, polypeptides and proteins can be purified according to standard procedures known to one of ordinary skill in the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis, high performance liquid chromatography, immunoprecipitation and the like.  
10 Substantially pure compositions of about 50 to 99% homogeneity are preferred, and 80 to 95% or greater homogeneity are most preferred for use as therapeutic agents.

In one embodiment of the present invention, the animal in which the desired protein is produced is an egg-laying animal. In a preferred embodiment of the present invention, the animal is an avian and a desired peptide, polypeptide or protein is  
15 isolated from an egg white. Egg white containing the exogenous protein or peptide is separated from the yolk and other egg constituents on an industrial scale by any of a variety of methods known in the egg industry. See, e.g., W. Stadelman et al. (Eds.), Egg Science & Technology, Haworth Press, Binghamton, NY (1995). Isolation of the exogenous peptide or protein from the other egg white constituents is accomplished  
20 by any of a number of polypeptide isolation and purification methods well known to one of ordinary skill in the art. These techniques include, for example, chromatographic methods such as gel permeation, ion exchange, affinity separation, metal chelation, HPLC, hydrophobic interaction chromatography, and the like, either alone or in combination. Such techniques are known to one of ordinary skill in the art  
25 of protein purification. Another means that may be used for isolation or purification, either in lieu of or in addition to chromatographic separation methods, includes electrophoresis. Successful isolation and purification is confirmed by standard analytic techniques, including HPLC, mass spectroscopy, and spectrophotometry. These separation methods are often facilitated if the first step in the separation is the  
30 removal of the endogenous ovalbumin fraction of egg white, as doing so will reduce the total protein content to be further purified by about 50%.

To facilitate or enable purification of a desired protein or peptide, the polynucleotide cassettes may include one or more additional epitopes or domains.

Such epitopes or domains include DNA sequences encoding enzymatic, chemical or photolabile cleavage sites including, but not limited to, an enterokinase cleavage site; the glutathione binding domain from glutathione S-transferase; polylysine; hexahistidine or other cationic amino acids, and sites cleaved by cyanogen bromide, hydroxylamine, formic acid, and acetic acid; thioredoxin; hemagglutinin antigen; maltose binding protein; a fragment of gp41 from HIV; and other purification epitopes or domains commonly known to one of skill in the art. Other proteolytic cleavage sites that may be included in the polynucleotide cassettes are cleavage sites recognized by exopeptidases such as carboxypeptidase A, carboxypeptidase B, aminopeptidase I, and dipeptidylaminopeptidase; endopeptidases such as trypsin, V8-protease, enterokinase, factor Xa, collagenase, endoproteinase, subtilisin, and thrombin; and proteases such as Protease 3C IgA protease (Igase) Rhinovirus 3C(preScission)protease. Self-splicing cleavage sites such as inteins may also be included in the polynucleotide cassettes of the present invention.

In one representative embodiment, purification of desired proteins from egg white utilizes the antigenicity of the ovalbumin carrier protein and particular attributes of a TAG linker sequence that spans ovalbumin and the desired protein. The TAG sequence is particularly useful in this process because it contains 1) a highly antigenic epitope, a fragment of gp41 from HIV, allowing for stringent affinity purification, and, 2) a recognition site for the protease enterokinase immediately juxtaposed to the desired protein. In a preferred embodiment, the TAG sequence comprises approximately 50 amino acids. A representative TAG sequence is provided below.

Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp  
Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys  
Gly Trp Ile Gly Leu Leu Asp Asp Asp Asp Lys (SEQ ID NO:90)

The underlined sequences were taken from the hairpin loop domain of HIV gp-41 (SEQ ID NO:87). Sequences in italics represent the cleavage site for enterokinase (SEQ ID NO:89). The spacer sequence upstream of the loop domain was made from repeats of (Pro Ala Asp Asp Ala) (SEQ ID NO:85) to provide free rotation and promote surface availability of the hairpin loop from the ovalbumin carrier protein.

Isolation and purification of a desired protein is performed as follows:

1. Enrichment of the egg white protein fraction containing ovalbumin and the transgenic ovalbumin-TAG-desired protein.
2. Size exclusion chromatography to isolate only those proteins within a narrow range of molecular weights (a further enrichment of step 1).
- 5 3. Ovalbumin affinity chromatography. Highly specific antibodies to ovalbumin will eliminate virtually all extraneous egg white proteins except ovalbumin and the transgenic ovalbumin-TAG-desired protein.
4. gp41 affinity chromatography using anti-gp41 antibodies. Stringent application of this step will result in virtually pure transgenic ovalbumin-TAG-desired protein.
- 10 5. Cleavage of the transgene product can be accomplished in at least one of two ways:
  - a. The transgenic ovalbumin-TAG-desired protein is left attached to the gp41 affinity resin (beads) from step 4 and the protease enterokinase is added. This liberates the transgene target protein from the gp41 affinity resin while the ovalbumin-TAG sequence is retained. Separation by centrifugation (in a batch process) or flow through (in a column purification), leaves the desired protein together with enterokinase in solution. Enterokinase is recovered and reused.
  - 15 b. Alternatively, enterokinase is immobilized on resin (beads) by the addition of poly-lysine moieties to a non-catalytic area of the protease. The transgenic ovalbumin-TAG-desired protein eluted from the affinity column of step 4 is then applied to the protease resin. Protease action cleaves the ovalbumin-TAG sequence from the desired protein and leaves both entities in solution. The immobilized enterokinase resin is recharged and reused.
  - 20 c. The choice of these alternatives is made depending upon the size and chemical composition of the transgene target protein.
- 25 6. A final separation of either of these two (5a or 5b) protein mixtures is made using size exclusion, or enterokinase affinity chromatography. This step allows for desalting, buffer exchange and/or polishing, as needed.
- 30

Cleavage of the transgene product (ovalbumin-TAG-desired protein) by enterokinase, then, results in two products: ovalbumin-TAG and the desired protein.

More specific methods for isolation using the TAG label is provided in the Examples. Some desired proteins may require additions or modifications of the above-described approach as known to one of ordinary skill in the art. The method is scaleable from the laboratory bench to pilot and production facility largely because the techniques  
5 applied are well documented in each of these settings.

It is believed that a typical chicken egg produced by a transgenic animal of the present invention will contain at least 0.001 mg, from about 0.001 to 1.0 mg, or from about 0.001 to 100.0 mg of exogenous protein, peptide or polypeptide, in addition to  
10 the normal constituents of egg white (or possibly replacing a small fraction of the latter). One of skill in the art will recognize that after biological expression or purification, the desired proteins, fragments thereof and peptides may possess a conformation substantially different than the native conformations of the proteins, fragments thereof and peptides. In this case, it is often necessary to denature and reduce protein and then to cause the protein to re-fold into the preferred conformation.  
15 Methods of reducing and denaturing proteins and inducing re-folding are well known to those of skill in the art.

#### Production of Proteins and Multimeric Proteins in Milk

In addition to methods of producing eggs containing transgenic proteins or peptides, the present invention encompasses methods for the production of milk  
20 containing transgenic proteins or peptides. These methods include the administration of a transposon-based vector described above to a mammal, for example through administration into the mammary duct system in order to access epithelial cells. In another embodiment, the transposon-based vector may be administered into arteries supplying the mammary gland. In a further embodiment germline transformed  
25 animals may have the transposon-based vector stably incorporated into the mammary epithelial cells. In one embodiment, the transposon-based vector contains a transposase operably-linked to a constitutive promoter and a gene of interest operably-linked to mammary specific promoter. Genes of interest can include, but are not limited to antiviral and antibacterial proteins, immunoglobulins, antibodies and  
30 proteins associated with disease and infectious organisms. Milk may be processed for isolation of proteins, multimeric proteins and antibodies using techniques known to one of skill in the art of processing milk and purifying proteins and antibodies from milk.

Administration of the Vaccines Produced with the Present Invention

The proteins, polypeptides, peptides and antibodies made with the present invention for use in vaccines may be administered to animals or humans using techniques known to one of ordinary skill in the art. Clinical practitioners of ordinary skill in the in the veterinary and human medical fields are familiar with methods, routes, volumes and schedules of vaccine administration, as well as various adjuvants that may be optionally employed to engender or amplify an immune response.

The following examples will serve to further illustrate the present invention without, at the same time, however, constituting any limitation thereof. On the contrary, it is to be clearly understood that resort may be had to various embodiments, modifications and equivalents thereof which, after reading the description herein, may suggest themselves to those skilled in the art without departing from the spirit of the invention.

15

## EXAMPLE 1

*Preparation of Transposon-Based Vector pTnMod*

A vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells, given below as SEQ ID NO:57. The vector of SEQ ID NO:57, termed pTnMod, was constructed and its sequence verified.

20

This vector employed a cytomegalovirus (CMV) promoter. A modified Kozak sequence (ACCATG) (SEQ ID NO:5) was added to the promoter. The nucleotide in the wobble position in nucleotide triplet codons encoding the first 10 amino acids of transposase was changed to an adenine (A) or thymine (T), which did not alter the amino acid encoded by this codon. Two stop codons were added and a synthetic polyA was used to provide a strong termination sequence. This vector uses a promoter designed to be active soon after entering the cell (without any induction) to increase the likelihood of stable integration. The additional stop codons and synthetic polyA insures proper termination without read through to potential genes downstream.

25

30

The first step in constructing this vector was to modify the transposase to have the desired changes. Modifications to the transposase were accomplished with the primers High Efficiency forward primer (Hef) Altered transposase (ATS)-Hef 5' ATCTCGAGACCATGTGTGAACTTGATTTTACATGATTCTCTTTACC 3'



(SEQ ID NO:91) and Altered transposase- High efficiency reverse primer (Her) 5' GATTGATCATTATCATAATTTCCCAAAGCGTAACC 3' (SEQ ID NO:92, a reverse complement primer). In the 5' forward primer ATS-Hef, the sequence CTCGAG (SEQ ID NO:93) is the recognition site for the restriction enzyme Xho I, which permits directional cloning of the amplified gene. The sequence ACCATG (SEQ ID NO:5) contains the Kozak sequence and start codon for the transposase and the underlined bases represent changes in the wobble position to an A or T of codons for the first 10 amino acids (without changing the amino acid coded by the codon). Primer ATS-Her (SEQ ID NO:92) contains an additional stop codon TAA in addition to native stop codon TGA and adds a Bcl I restriction site, TGATCA (SEQ ID NO:94), to allow directional cloning. These primers were used in a PCR reaction with pTnLac (p defines plasmid, tn defines transposon, and lac defines the beta fragment of the lactose gene, which contains a multiple cloning site) as the template for the transposase and a FailSafe<sup>TM</sup> PCR System (which includes enzyme, buffers, dNTP's, MgCl<sub>2</sub> and PCR Enhancer; Epicentre Technologies, Madison, WI). Amplified PCR product was electrophoresed on a 1% agarose gel, stained with ethidium bromide, and visualized on an ultraviolet transilluminator. A band corresponding to the expected size was excised from the gel and purified from the agarose using a Zymo Clean Gel Recovery Kit (Zymo Research, Orange, CA). Purified DNA was digested with restriction enzymes Xho I (5') and Bcl I (3') (New England Biolabs, Beverly, MA) according to the manufacturer's protocol. Digested DNA was purified from restriction enzymes using a Zymo DNA Clean and Concentrator kit (Zymo Research).

Plasmid gWhiz (Gene Therapy Systems, San Diego, CA) was digested with restriction enzymes Sal I and BamH I (New England Biolabs), which are compatible with Xho I and Bcl I, but destroy the restriction sites. Digested gWhiz was separated on an agarose gel, the desired band excised and purified as described above. Cutting the vector in this manner facilitated directional cloning of the modified transposase (mATS) between the CMV promoter and synthetic polyA.

To insert the mATS between the CMV promoter and synthetic polyA in gWhiz, a Stratagene T4 Ligase Kit (Stratagene, Inc. La Jolla, CA) was used and the ligation set up according to the manufacturer's protocol. Ligated product was transformed into *E. coli* Top10 competent cells (Invitrogen Life Technologies,

Carlsbad, CA) using chemical transformation according to Invitrogen's protocol. Transformed bacteria were incubated in 1 ml of SOC (GIBCO BRL, CAT# 15544-042) medium for 1 hour at 37° C before being spread to LB (Luria-Bertani media (broth or agar)) plates supplemented with 100 µg/ml ampicillin (LB/amp plates).  
5 These plates were incubated overnight at 37° C and resulting colonies picked to LB/amp broth for overnight growth at 37° C. Plasmid DNA was isolated using a modified alkaline lysis protocol (Sambrook et al., 1989), electrophoresed on a 1% agarose gel, and visualized on a U.V. transilluminator after ethidium bromide staining. Colonies producing a plasmid of the expected size (approximately 6.4 kbp)  
10 were cultured in at least 250 ml of LB/amp broth and plasmid DNA harvested using a Qiagen Maxi-Prep Kit (column purification) according to the manufacturer's protocol (Qiagen, Inc., Chatsworth, CA). Column purified DNA was used as template for sequencing to verify the changes made in the transposase were the desired changes and no further changes or mutations occurred due to PCR amplification. For  
15 sequencing, Perkin-Elmer's Big Dye Sequencing Kit was used. All samples were sent to the Gene Probes and Expression Laboratory (LSU School of Veterinary Medicine) for sequencing on a Perkin-Elmer Model 377 Automated Sequencer.

Once a clone was identified that contained the desired mATS in the correct orientation, primers CMVf-NgoM IV (5' TTGCCGGCATCAGATTGGCTAT (SEQ  
20 ID NO:95); underlined bases denote NgoM IV recognition site) and Syn-polyA-BstE II (5' AGAGGTCACCGGGTCAATTCTTCAGCACCTGGTA (SEQ ID NO:96); underlined bases denote BstE II recognition site) were used to PCR amplify the entire CMV promoter, mATS, and synthetic polyA for cloning upstream of the transposon in pTnLac. The PCR was conducted with FailSafe™ as described above, purified  
25 using the Zymo Clean and Concentrator kit, the ends digested with NgoM IV and BstE II (New England Biolabs), purified with the Zymo kit again and cloned upstream of the transposon in pTnLac as described below.

Plasmid pTnLac was digested with NgoM IV and BstE II to remove the ptac promoter and transposase and the fragments separated on an agarose gel. The band  
30 corresponding to the vector and transposon was excised, purified from the agarose, and dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs) to prevent self-annealing. The enzyme was removed from the vector using a Zymo DNA Clean and Concentrator-5. The purified vector and CMVp/mATS/polyA

were ligated together using a Stratagene T4 Ligase Kit and transformed into *E. coli* as described above.

Colonies resulting from this transformation were screened (mini-preps) as describe above and clones that were the correct size were verified by DNA sequence analysis as described above. The vector was given the name pTnMod (SEQ ID NO:57) and includes the following components:

Base pairs 1-130 are a remainder of F1(-) on from pBluescriptII sk(-) (Stratagene), corresponding to base pairs 1-130 of pBluescriptII sk(-).

Base pairs 131 - 132 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 133 -1777 are the CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems), corresponding to bp 229-1873 of pGWiz. The CMV promoter was modified by the addition of an ACC sequence upstream of ATG.

Base pairs 1778-1779 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 1780 - 2987 are the coding sequence for the transposase, modified from Tn10 (GenBank accession J01829) by optimizing codons for stability of the transposase mRNA and for the expression of protein. More specifically, in each of the codons for the first ten amino acids of the transposase, G or C was changed to A or T when such a substitution would not alter the amino acid that was encoded.

Base pairs 2988-2993 are two engineered stop codons.

Base pair 2994 is a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 2995 - 3410 are a synthetic polyA sequence taken from the pGWiz vector (Gene Therapy Systems), corresponding to bp 1922-2337 of 10 pGWiz.

Base pairs 3415 - 3718 are non-coding DNA that is residual from vector pNK2859.

Base pairs 3719 - 3761 are non-coding  $\lambda$  DNA that is residual from pNK2859.

Base pairs 3762 - 3831 are the 70 bp of the left insertion sequence recognized by the transposon Tn10.

Base pairs 3832-3837 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 3838 - 4527 are the multiple cloning site from pBluescriptII sk(20), corresponding to bp 924-235 of pBluescriptII sk(-). This multiple cloning site may be used to insert any coding sequence of interest into the vector.

Base pairs 4528-4532 are a residue from ligation of restriction enzyme sites  
5 used in constructing the vector.

Base pairs 4533 - 4602 are the 70 bp of the right insertion sequence recognized by the transposon Tn10.

Base pairs 4603 - 4644 are non-coding  $\lambda$  DNA that is residual from pNK2859.

Base pairs 4645 - 5488 are non-coding DNA that is residual from pNK2859.

Base pairs 5489 - 7689 are from the pBluescriptII sk(-) base vector -  
10 (Stratagene, Inc.), corresponding to bp 761-2961 of pBluescriptII sk(-).

Completing pTnMod is a pBlueScript backbone that contains a colE I origin of replication and an antibiotic resistance marker (ampicillin).

It should be noted that all non-coding DNA sequences described above can be  
15 replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

All plasmid DNA was isolated by standard procedures. Briefly, *Escherichia coli* containing the plasmid was grown in 500 mL aliquots of LB broth (supplemented with an appropriate antibiotic) at 37°C overnight with shaking. Plasmid DNA was  
20 recovered from the bacteria using a Qiagen Maxi-Prep kit (Qiagen, Inc., Chatsworth, CA) according to the manufacturer's protocol. Plasmid DNA was resuspended in 500  $\mu$ L of PCR-grade water and stored at -20°C until used.

## EXAMPLE 2

### 25 *Transposon-Based Vector pTnMCS*

Another transposon-based vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells. This vector was termed pTnMCS and its constituents are provided below. The sequence of the pTnMCS vector is provided in SEQ ID NO:56. The pTnMCS vector contains an avian optimized polyA sequence  
30 operably-linked to the transposase gene. The avian optimized polyA sequence contains approximately 75 nucleotides that precede the A nucleotide string.

Bp 1 - 130 Remainder of F1 (-) ori of pBluescriptII sk(-) (Stratagene) bp1-130

- Bp 133 – 1777 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems) bp 229-1873
- Bp 1783 – 2991 Transposase, from Tn10 (GenBank accession #J01829) bp 108-1316
- Bp 2992 – 3344 Non coding DNA from vector pNK2859
- 5 Bp 3345 – 3387 Lambda DNA from pNK2859
- Bp 3388 – 3457 70 bp of IS10 left from Tn10
- Bp 3464 – 3670 Multiple cloning site from pBluescriptII sk(-), thru the XmaI site bp 924-718
- Bp 3671 - 3715 Multiple cloning site from pBluescriptII sk(-), from the XmaI site
- 10 thru the XhoI site. These base pairs are usually lost when cloning into pTnMCS bp 717-673
- Bp 3716 – 4153 Multiple cloning site from pBluescriptII sk(-), from the XhoI site bp 672-235
- Bp 4159 - 4228 70 bp of IS10 right from Tn10
- 15 Bp 4229 - 4270 Lambda DNA from pNK2859
- Bp 4271 – 5114 Non-coding DNA from pNK2859
- Bp 5115 - 7315 pBluescript sk (-) base vector (Stratagene, Inc.) bp 761-2961.

### EXAMPLE 3

#### 20 *Production of Antibody in Egg White*

A transposon-based vector containing a CMV promoter/cecropin prepro/antibody heavy chain/cecropin pro/Antibody light chain/conalbumin poly A (SEQ ID NO:97) was injected into the oviduct of quail and chickens. A total of 20 birds were injected (10 chickens and 10 quail) and eggs were harvested from the birds

25 once the eggs were laid. Partially purified egg white protein (EW) was then run on a gel under both reducing and non-reducing conditions. Figure 5 is a picture of the gel. Lanes 1 & 18: molecular weight markers, Lanes 2 and 3: EW #1, non-reduced, reduced, respectively; Lanes 4 and 5: EW #2, non-reduced, reduced, respectively, Lanes 6 and 7: EW #3, non-reduced, reduced, respectively, Lanes 8 and 9: EW #4,

30 non-reduced, reduced, respectively; Lanes 10 and 11: EW #5, non-reduced, reduced, respectively; Lanes 12 and 13: EW #6, non-reduced, reduced, respectively; Lanes 14 and 15: EW #7, non-reduced, reduced, respectively; and Lanes 16 and 17: EW #8 Control, non-reduced, reduced, respectively. Based upon the gel results, the

possibility that the egg white in the treated chicken and quail contains antibody produced by the above-mentioned transposon-based vector cannot be excluded.

#### EXAMPLE 4

##### 5 *Additional Transposon-Based Vectors for Administration to an Animal*

The following example provides a description of various transposon-based vectors of the present invention and several constructs for insertion into the transposon-based vectors of the present invention, all for intraoviduct administration. These examples are not meant to be limiting in any way. The constructs for insertion  
10 into a transposon-based vector are provided in a cloning vector pTnMCS or pTnMod, both described above.

##### pTnMOD (CMV-prepro-HCPro-Lys-CPA) (SEQ ID NO:97)

Bp 1-4090 from vector pTnMod, bp 1 - 4090

Bp 4096-5739 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy  
15 systems), bp 230-1864

Bp 5746-5916 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733

Bp 5923-7287 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7288-7302 Pro taken from GenBank accession # X07404, bp 719-733 (includes  
20 Lysine)

Bp 7309-7953 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7960-8372 Conalbumin polyA taken from GenBank accession # Y00407, bp  
10651-11058

25 Bp 8374-11973 from cloning vector pTnMod, bp 4091-7690

##### pTnMCS (CHOVep-prepro-HCPro-CPA) (SEQ ID NO:98)

Bp 1-3715 from vector pTnMCS, bp 1-3715

Bp 3721-4395 Chicken Ovalbumin enhancer taken from GenBank accession #  
30 S82527.1, bp 1-675

Bp 4402-5738 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336

Bp 5745-5915 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733

- Bp 5922-7286 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
- Bp 7287-7298 Pro taken from GenBank accession # X07404, bp 719-730 (does not include Lysine)
- 5 Bp 7305-7949 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
- Bp 7956-8363 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
- Bp 8365-11964 from cloning vector pTnMCS, bp 3716-7315
- 10 pTnMCS(CHO<sub>vep</sub>-prepro-HCPro-Lys-CPA) (SEQ ID NO:99)
- Bp 1 – 3715 from vector pTnMCS, bp 1-3715
- Bp 3721 – 4395 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1-675
- 15 Bp 4402 - 5738 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336
- Bp 5745 – 5915 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733
- Bp 5922 - 7286 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
- 20 Bp 7287 - 7301 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)
- Bp 7308 – 7952 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
- Bp 7959 – 8366 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
- 25 Bp 8368 – 11967 from cloning vector pTnMCS, bp 3716-7315
- pTnMCS (CMV-prepro-HCPro-CPA) (SEQ ID NO:100)
- Bp 1 – 3715 from vector pTnMCS, bp 1-3715
- 30 Bp 3721 – 5364 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864
- Bp 5371-5541 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733

- Bp 5548 - 6912 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
- Bp 6913 - 6924 Pro taken from GenBank accession # X07404, bp 719-730 (does not Lysine)
- 5 Bp 6931 - 7575 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
- Bp 7582 - 7989 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
- Bp 7991 - 11590 from cloning vector pTnMCS, bp 3716-7315
- 10 pTnMCS (CMV-prepro-HCPro-Lys-CPA) (SEQ ID NO:101)
- Bp 1 - 3715 from vector pTnMCS, bp 1-3715
- Bp 3721 - 5364 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864
- 15 Bp 5371-5541 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733
- Bp 5548 - 6912 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
- Bp 6913 - 6927 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)
- 20 Bp 6934 - 7578 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
- Bp 7585 - 7992 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
- Bp 7994 - 11593 from cloning vector pTnMCS, bp 3716-7315
- 25 pTnMod (CHO<sub>vep</sub>-prepro-HCPro-CPA) (SEQ ID NO:102)
- Bp 1-4090 from vector pTnMod, bp 1-4090
- Bp 4096-4770 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1-675
- 30 Bp 4777-6113 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336
- Bp 6120-6290 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733



Bp 6297-7661 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7662-7673 Pro taken from GenBank accession # X07404, bp 719-730 (does not include Lysine)

5 Bp 7680-8324 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 8331-8738 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 8740-12339 from cloning vector pTnMod, bp 3716-7315

10

pTnMod (CHO<sub>vep</sub>-prepro-HCPro-LYS-CPA) (SEQ ID NO:103)

Bp 1-4090 from vector pTnMod, bp 1-4090

Bp 4096-4770 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1-675

15 Bp 4777-6113 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336

Bp 6120-6290 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733

Bp 6297-7661 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

20 Bp 7662-7676 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)

Bp 7683-8327 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 8334-8741 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

25

Bp 8743-12342 from cloning vector pTnMod, bp 3716-7315

pTnMod (CMV-prepro-HCPro-CPA) (SEQ ID NO:104)

Bp 1-4090 from vector pTnMod, bp 1-4090

30 Bp 4096-5739 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864

Bp 5746-5916 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733

Bp 5923-7287 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7288-7299 Pro taken from GenBank accession # X07404, bp 719-730 (does not include Lysine)

5 Bp 7306-7950 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7557-7969 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 7971-11970 from cloning vector pTnMod, bp 3716-7315

10

#### Example 5

*Production of a Vaccine in Transgenic Chickens against Brucella Melitensis, Collection of the Vaccine from Eggs and Use of the Vaccine to Confer a Protective Response*

15 *Brucella* species are considered potential biological warfare agents due to the characteristic protracted debilitating disease caused by these organisms. The antigens of interest are the Group 3 Omps which have been shown to demonstrate a role in protective immunity. Group 3 Omps are comprised of two separate proteins of 25 kDa (Omp25) and 31 kDa (Omp31) (Cloeckaert et al., 1996a).

20 First, genes encoding *Brucella melitensis* outer membrane proteins (Omp 25 and Omp 31) were cloned for expression into egg albumin. The vectors that are used have been described above and the genes of interest encode for Omp 25 and Omp 31.

Harvest of *Brucella* Genomic DNA: Genomic DNA from *B. melitensis* is harvested from bacteria grown in 250 ml of *Brucella* Broth (Difco Laboratories, Detroit, MI) for 18 hours at 37° C. The culture is killed by the addition of chloroform for 1 h to prevent accidental infection of lab personnel. After killing, genomic DNA isharvested using a Qiagen Genomic-Tip System (Qiagen, Inc., Valencia, CA) according to the manufacturer's protocol and resuspended in PCR grade water.

25 PCR Optimization: For PCR optimization, primers specific to Omp25 and Omp31 are used to amplify each gene from genomic DNA of *B. melitensis*. Conditions for PCR with each set of primers are optimized using a FailSafe™ PCR System kit (Epicentre Technologies, Madison, WI), the PCR fragments separated by agarose gel electrophoresis, excised from the gel, and purified from the agarose using

a QIAquick Gel Extraction Kit (Qiagen, Inc). Purified DNA corresponding to Omp25 and Omp31 is stored at  $-20^{\circ}\text{C}$  until needed.

5 Plasmid DNA Isolation: All plasmid DNA is isolated by procedures routinely used in the laboratory of Dr. Cooper. Briefly, *Escherichia coli* containing the desired plasmid is grown in 500 ml aliquots of LB broth (supplemented with the appropriate antibiotic) at  $37^{\circ}\text{C}$  overnight with shaking. Plasmid DNA is recovered from the bacteria using a Qiagen Maxi-Prep kit (Qiagen, Inc.) according to the manufacturer's protocol. Plasmid DNA is resuspended in  $500\ \mu\text{L}$  of PCR grade water and stored at  $20^{\circ}\text{C}$  until needed.

10 The vectors are digested with restriction enzymes. Bands corresponding to the vector are excised from the gel, purified from agarose (see above), and ligated (Fast-Link<sup>TM</sup> DNA Ligation Kit, Epicentre Technologies) overnight with either DNA encoding Omp25 or Omp31. Ligation mixtures are electroporated into *E. coli* STBL4<sup>TM</sup> (Life Technologies, Rockville, MD), plated onto LB agar plates  
15 supplemented with  $200\ \mu\text{g/ml}$  ampicillin, and incubated at  $37^{\circ}\text{C}$  overnight.

Resulting colonies are grown at  $37^{\circ}\text{C}$  overnight in 2 ml of LB broth supplemented with  $200\ \mu\text{g/ml}$  ampicillin, plasmid DNA harvested using a QIAprep Spin Miniprep Kit (Qiagen, Inc.), and separated by agarose gel electrophoresis. Bacterial isolates containing the expected insert size are frozen at  $-80^{\circ}\text{C}$ . The insert  
20 is sequenced to determine orientation and verify the sequence. All sequencing is conducted using a BigDye<sup>TM</sup> Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, Foster City, CA) according to the manufacturer's protocol and loaded onto a PE Applied Biosystems 310 Genetic Analyzer. Isolates containing Omp25 (pTnOv25) and Omp31 (pTnOv31) with the correct sequence and orientation  
25 are grown as described above for Maxi-Prep DNA isolation. Purified DNA will be resuspended in TE buffer ( $10\ \text{mM}$  tris,  $5\ \text{mM}$  EDTA, pH 8.0) and stored at  $-20^{\circ}\text{C}$  until needed.

Next transgenic chickens are developed to produce Omp25 and Omp31 in the albumin portion of their eggs. The experimental males are obtained from commercial  
30 sources at one day of age and maintained in brooders until used. After being placed in a treatment group, males will be housed individually in standard caging in temperature controlled spaces appropriate to the age of the birds. They are watered and fed standard Starter Chick chow *ad lib* and kept initially in a 23:1 h light/dark

cycle, stepping down at approximately weekly intervals to a 15h:8h light/dark cycle, as this regimen optimizes sexual maturity and fertility (Duncan 1999).

For transfection, plasmid DNA is complexed with Superfect™ (Qiagen, Inc) in an approximate 1:3 ratio, following the manufacturer's protocol. The DNA is injected (described below) at a rate of 5 µg per testis in a 50 µL volume [Hank's balanced salt solution (HBSS) will be used to bring the total volume to 50 µl]. At least 20 males are used for controls and receive an injection of HBSS and transfecting reagent. All birds are held until sexual maturity before sperm samples are obtained for PCR.

At age 30-98 days, groups of individual males are subjected to transgene delivery by direct intratesticular injection of DNA. Males are anesthetized with isoflourane via a gas machine and positioned on their side on an animal board with cords tractioning wings and feet to allow access to the testes area. The area is swabbed with 0.5% chlorhexidine and a 2 cm dorsolateral incision made over the testis (which is identical to the procedure for caponization). After exposing the testis, a capillary injection device (described below) is used to inject, in 2 or 3 sites, about 50 µl of the DNA/SuperFect complex (Qiagen) (~5 µg DNA per testis). The injections are calculated to suffuse the DNA throughout the whole testis, allowing as much contact of DNA with spermatogonia as possible. Following injection, the wound is closed in two layers with 4-0 absorbable suture, and the contralateral testis similarly exposed and injected. Following surgery, the bird is returned to its cage to recover. Approximately 50 males are used for this experimental regimen to increase the likelihood of obtaining a male having transgenic sperm.

The injection device consists of a 27 ga hypodermic needle cut off at the hub, bent to the desired angle, and welded onto a standard 100 µl capillary tube. The capillary is filled with the DNA solution and placed into a standard mouth pipette so that the injection rate and volume can be accurately controlled hands-free. After injection, males are allowed to reach sexual maturity in order to obtain sperm for PCR. Males with sperm positive for Omp25 or Omp31 are mated with females to produce transgenic F1s. The F1s are verified for copy number of the transgene. These transgenic F1s are raised to maturity.

Next the expressed proteins are purified from the eggs of the transgenic chickens and subsequently used in vaccine efficacy trials. The purification

is achieved with methods presented above for purification of proteins from eggs. The following is only one approach to purify the protein. The initial objective in the purification of expressed proteins in egg white is to eliminate the major component of egg white, namely, the albumin proteins. Li – Chan, et al., 1986 have described a  
5 versatile and simple ion – exchange procedure which largely accomplishes this elimination of the albumin.

The egg white is lightly homogenized to reduce viscosity [shear rate of 3140  $\text{sec}^{-1}$  reduces albumin viscosity from 21 to 5 centipoises (Tung et al., 1969)]. The albumin solution is filtered through two layers of muslin to remove chalazae. The  
10 resulting solution is then passed through an industrial cation – exchanger, IMAC-HP333 (Rohm and Haas, Philadelphia, PA) which has been equilibrated at pH 9 with low ionic strength buffer (phosphate or bicarbonate). The albumin passes through the ion – exchange column leaving some ovomucin and all of the avidin and lysozyme on the column.

15 This procedure allows cationic proteins/peptides to be separated from the albumin at an early stage in the purification. What is required for making this process a general one is to either purify such proteins which are by their nature, highly positively charged (e.g. cecropin – like lytic peptides or their derivatives) or utilize fusion proteins wherein the fusion adduct contains the requisite cationic charge. Also  
20 required is that the magnitude of charge on the expressed protein or fusion protein must exceed considerably that of any possible cationic contaminant.

Once the cationic entities have been captured by the ion – exchanger, elution is initiated by a series of increasing ionic strength, pH 8.0 bicarbonate buffers. The desired product is retained until it is selectively eluted at the highest ionic strength  
25 (ammonium bicarbonate, approaching 2M). The water and most of the ammonium bicarbonate are removed by lyophilization. The recovered fusion protein is dissolved in 88% formic acid and subjected to cyanogen bromide cleavage. If the desired component is anionic, the lyophilized residue is dissolved in dilute acetic acid and passed through and anion – exchange resin contained in a jacketed column. The  
30 column is washed with dilute hydrochloric acid (0.05 M) and cold ( $-5^{\circ}\text{C}$ ) ethylene glycol solution pumped through the jacket on the column. By-products from the cleavage reaction are substantially removed by this step. The anionic protein is then lyophilized (using solid KOH in the vacuum system to capture the HCl) to dryness,

dissolved in a minimal amount of water and lyophilized (twice) to remove residual HCl.

If the desired component of the fusion protein is cationic then the cleavage mixture is diluted 5 – fold with acetate buffer and passed onto a sulfoethylcellulose cation exchanger contained in a jacketed column. The column is washed with 0.05 M HCl and then 0.15 M HCl to remove residual cationic materials including degradation products while pumping cold (-5°C) ethylene glycol through the column jacket. The desired product is eluted with 1.0 M HCL adjusted carefully to pH 2.7 by the extremely slow addition of very cold ammonium hydroxide. The eluted protein is lyophilized to dryness using solid KOH within the vacuum system to capture the HCl. The protein is dissolved in a minimum amount of water and lyophilized (twice) to remove residual HCl. Protein quantity will be determined using a Bio-Rad Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA). Purity is determined with Western blots, HPLC, ELISAs and other methods known to one of ordinary skill in the art.

Vaccination and Challenge of Mice: The BALB/c murine model of brucellosis established by Montaraz and Winter (1986) has provided much of the current understanding of immunity to brucellosis. The model is valuable since the murine colonization profile of virulent type strains and vaccine candidates mimics what is often attenuated in the ruminant host (Montaraz and Winter, 1986; Winter, 1990; Schurig et al., 1991; Elzer et al., 1996; Stevens et al., 1997; Edmonds et al., 2000a; Edmonds et al., 2000b). The murine model is also useful due to practical concerns including availability of reagents, cost, and population homogeneity (Montaraz and Winter, 1986; Garcia-Carrillo, 1990).

Bacterial Strains: *Brucella melitensis* strain 16M has been documented to cause chronic infections in mice (Jiminez et al., 1994; Philips et al., 1995). Infectious doses containing approximately  $1 \times 10^9$  and  $1 \times 10^{11}$  CFU/ml are prepared as described (Elzer et al., 1994c) and snap frozen in liquid nitrogen prior to storage at -80°C. Immediately prior to animal inoculation, samples are thawed and diluted in sterile saline to the appropriate concentration. Doses are verified by serial dilution and plating on Schaedler agar containing 5% bovine blood (SBA) (Difco Laboratories, Detroit, MI).

Female BALB/c mice at approximately 6 to 8 weeks of age are obtained from the Division of Laboratory Animal Medicine at the School of Veterinary Medicine at Louisiana State University (Baton Rouge, LA) and held at least one week prior to use. All animals are housed in a restricted access facility. All mice are used in accordance with the Animal Care and Use Committee at Louisiana State University, and approved protocols will be followed.

Murine Vaccine Efficacy Trials: Mice are divided into groups of ten and are given the following vaccines intraperitoneal (i.p.): Group 1 – saline control; Group 2 –  $5 \times 10^4$  CFU of *B. melitensis* vaccine strain Rev-1 i.v.; Group 3 – Titremax adjuvant; Group 4 – Omp25 plus Titremax adjuvant; Group 5 – Omp31 plus Titremax adjuvant; and Group 6 – Omp25 and Omp31 plus Titremax adjuvant. Mice will be vaccinated with 30  $\mu$ g *B. melitensis* Omp25 or Omp31. The dose of the Rev-1 strain is based on previously published work in BALB/c mice (Jimenez de Bagues et al., 1994a; Jimenez de Bagues et al., 1994b). Prior to vaccination and 8 weeks post-vaccination, mice were bled and the serum pooled for each group and the Omp25 and Omp31 vaccinated mice are boosted.

Ten weeks following initial vaccination with either Omp25, Omp31, Omp25 plus Omp31, or *B. melitensis* strain Rev. 1, both vaccinated and non-vaccinated mice are i.v. inoculated with the challenge strain  $5 \times 10^4$  CFU of *B. melitensis* strain 16M (Jimenez de Bagues et al., 1994a; Jimenez de Bagues et al., 1994b). Two weeks post-challenge, all mice are euthanized by halothane overdose. At necropsy, the whole spleen is removed by aseptic technique for bacteriological analysis while blood is obtained for serological analysis. Spleens are homogenized with a Sorvall Omni-Mixer (Newton, CT) in sterile phosphate-buffered-saline (PBS) and subsequently serially diluted and plated on SBA. All plates are incubated at 37°C in a 5% CO<sub>2</sub> atmosphere for 3 to 4 days until isolated colonies appear.

Serological procedures: Blood is collected from all the mice before inoculation with *B. melitensis* and at two-week intervals thereafter until they are killed. The sera obtained is stored at -80°C. *Brucella*-specific antibodies are detected by the Card test and Western blot analysis with whole-cell lysates of *B. melitensis* (Elzer et al., 1994a).

### Vaccination and challenge of Goats: *Strains*

- B. melitensis* is of established virulence in pregnant goats (Meador and Deyoe 1986, Enright 1990). The construction and phenotypic characteristics of the isogenic 2308 *htrA* mutant, PHE1, have been described by Elzer et al., (1994a). Infectious doses containing approximately  $10^9$  or  $10^{11}$  colony forming units (cfu)  $\text{ml}^{-1}$  in *Brucella* broth (Difco, Detroit, MI, USA) were prepared as described by Elzer et al., (1994a) and snap frozen in a dry ice-ethanol bath before being stored at  $-80^\circ\text{C}$ . These preparations were thawed just before use and their content of viable brucellae was assessed by serial dilution and plating.
- 10 Serological procedures: Blood is collected from all the goats before they are vaccinated, and at two-week intervals thereafter until they are killed. The sera obtained is stored at  $-80^\circ\text{C}$ . *Brucella*-specific antibodies are detected by the Card test and Western blot analysis with whole-cell lysates of *B abortus* (Elzer et al., 1994a).
- 15 *Vaccine Efficacy Study:* Seventy-two mature, non-pregnant female goats are divided into 6 groups and given 100  $\mu\text{g}$  of the following vaccines subcutaneously (s.c.): Group 1 – saline control; Group 2 –  $5 \times 10^4$  CFU of *B. melitensis* vaccine strain Rev-1 i.v.(this group receives saline first and then Rev-1 2 weeks later); Group 3 – Titremax adjuvant; Group 4 – Omp25 plus Titremax adjuvant; Group 5 – Omp31 plus Titremax
- 20 adjuvant; and Group 6 – Omp25 and Omp31 plus Titremax adjuvant. Eight weeks later, the 3 Omp groups and adjuvant control group are boosted with a second dose of the appropriate formulation. Twelve other animals are vaccinated SC with  $1 \times 10^6$  CFU of the vaccine *B. melitensis* strain Rev. 1. Previous experiments in goats have shown that this dose of strain Rev. 1 provides protection against challenge with
- 25 virulent *B. melitensis* strain 16M (Elzer et al., 1998). The remaining 12 animals serve as controls and receive another SC inoculation of the saline diluent. Four weeks after vaccination, a fertile male goat is housed with each group. Following ultrasound confirmation of pregnancy, all animals are conjunctively challenged at approximately 110 days gestation with  $1 \times 10^7$  CFU of virulent *B. melitensis* strain 16M. Following
- 30 inoculation, animals are monitored daily for delivery. At parturition, the birth status of the kid is recorded and within 24 h, necropsy samples obtained for bacteriologic analysis from the lung and abomasal fluid. Within one month of delivery, dams are



5 euthanized and samples taken for bacteriologic analysis from the parotid lymph node, prescapular lymph node, supramammary lymph node, internal iliac lymph node, liver, spleen, milk, and uterine fluid. Tissue samples are frozen at -4°C for later analysis; milk and uterine fluid are swabbed the day of collection. Serum samples are obtained prior to vaccination, at challenge, and at necropsy for serological analysis.

10 Bacteriological procedures: All the tissue specimens are homogenized in sterile PBS (0.002M NaH<sub>2</sub>PO<sub>4</sub>, 0.008M Na<sub>2</sub>HPO<sub>4</sub>, 0.15M NaCl, pH 7.2), plated on to Farrell's selective medium (Farrell 1974) and Schaedler agar supplemented with 5% sheep blood. The plates are incubated for 14 days at 37°C in an atmosphere of 5% carbon dioxide. The isolates were identified as *Brucella* species by their Gram-staining characteristics, colonial morphology, urease and oxidase reactions and agglutination with *B. abortus*-specific antibodies in a standard slide agglutination assay. (Elzer et al., 1994a)

15 Mice Studies: Mice vaccinated with the subcellular components (OMP) in conjunction with the adjuvant Titermax exhibit increased protective immunity (i.e., lower CFU/gm of tissue) as compared to the saline controls. Serological analysis remains negative since the *Brucella* diagnostic tests are dependent on antibody reactions to the O-side chain of the organism. The adjuvant alone does not contribute to the immune response or interfere with the serological analyses. The combination of the Titermax and OMP vaccines results in a predominant TH1 protective immune response as demonstrated by a lower bacterial load in cultured tissues.

25 Goat Studies: Goats receiving the Titermax and OMP combinations exhibit a protective immune response post-challenge as demonstrated by decreased colonization of the dam/fetus pairs and/or decreased abortions as compared to the saline or Titermax alone controls. *B. melitensis* Rev 1 is the accepted vaccine for caprine brucellosis; it is the "gold standard" for protection in this model. The Titermax/OMP vaccinations results in equivalent or better protection as compared to Rev 1. However, unlike Rev 1, the Titermax/OMP groups remain serologically negative on all diagnostic tests until after challenge when all groups become positive.

All patents, publications and abstracts cited above are incorporated herein by reference in their entirety. It should be understood that the foregoing relates only to preferred embodiments of the present invention and that numerous modifications or alterations may be made therein without departing from the spirit and the scope of the present invention as defined in the following claims.

5

## Appendix A

- 5 SEQ ID NO:1 (cecropin pro)  
GCG CCA GAG CCG AAA
- 10 SEQ ID NO:2 (cecropin pro extended)  
GCG CCA GAG CCG AAA TGG AAA GTC TTC AAG
- 15 SEQ ID NO:3 (cecropin prepro)  
AAT TTC TCA AGG ATA TTT TTC GTG TTC GCT TTG GTT CTG GCT TTG TCA ACA  
GTT TCG GCT GCG CCA GAG CCG AAA
- 20 SEQ ID NO:4 (cecropin prepro extended)  
AAT TTC TCA AGG ATA TTT TTC GTG TTC GCT TTG GTT CTG GCT TTG TCA ACA  
GTT TCG GCT GCG CCA GAG CCG AAA TGG AAA GTC TTC AAG
- 25 SEQ ID NO:5 (modified Kozak sequence)  
ACCATG
- 30 SEQ ID NO:6  
atg ctg ggc atc tgg acc ctc cta cct ctg gtt ctt acg tct gtt gct aga tta
- 35 SEQ ID NO:7  
atg att cct gcc aga ttt gcc ggg gtg ctg ctt gct ctg gcc ctc att ttg cca  
ggg acc ctt tgt
- 40 SEQ ID NO:8  
atgg gcagagcaat ggtggccagg ctggggctgg ggctgtgct gctgggcactg ctctaccca  
cgcagattta ttcc
- 45 SEQ ID NO:9  
atgaatttat cgaacatttc tgcggtaaaa qtattaacac tgggtggttag cgctgccatc gct

SEQ ID NO:10

atgaccatcc ttttccttac tatggttacc tcatacttca gttgcattgaa agtgcctccg  
 atgaaagaag ctagtgtgag aggacatggc agcttggctt accagggtctc tcggacccac  
 gggactcttg aaagcctaac tgggccaat gctggttcaa gaggactgac atcactggcg  
 gacacttttg aacacgtgat agaggagctt ctatgataag atcaggacat ccagccagct  
 gaggaaaaa aggatgcgga ctgtacaca tcccgagtca tgcagagcag tcaagtgcct  
 ttggaacccc cactgctctt tctgctcgag gagtacaaaa actacctgga tgcctgcaaac  
 atgtccatga gagtccggcg tcactctgac

SEQ ID NO:11

cgtctttttc tcttatcttt tctcgctttc gctcttttct cgtcggcgat tgccttctcc  
 gacgacgac cgttgatccg acaagttgta tcgggaaacg atgacaacca tatgttaaac  
 gccgagcacc acttttccact tttt

SEQ ID NO:12

atg tccatcttgt tttatgtgat atttcttga tatctctgtg gcattcagtc aactaatatg  
 gatcaaggga gtttgccaga agattcaatg aattctctca ttattaaact cattcgggca  
 gacatcttga aaacaagct ttctaagcag gtgatggatg tcaaggaaaa ctatcaaaac  
 atagtgcaga aagtagagga ccaccaggag atggatggag atgaaaaatgt gaaatcagac  
 ttccagccag ttatttcaat ggatacagac ctctaaggc agcagagacg ctacaactct  
 cctcgagttc tctaagtga caacacacca ctggaaccac caccactgta cctcacagag  
 gattatgtg gaagttcagt ggtattaaac agaactcttc gaaggaaaaag gt

SEQ ID NO:13

atggtgcacc tgactcctga ggagaagtct gccgttactg cctgtgggg caaggtgaac  
 gtggatgaag ttggtggtga ggcctgggc aggtgctgg tggctaccc ttggacccag  
 aggttctttg agtcccttgg gatatctgcc actcctgatg ctgttatgg caaccctaaag  
 gtgaaggctc atggcaagaa agtgcctgggt gcctttagt atggcctggc tcacctggac  
 aacctcaagg gcacctttgc cacactgagt gagctgcact gtgacaagct gcacgtggat  
 cctgagaact tcaggctcct gggcaacgtg ctggtctgtg tgcctggcca tcactttggc  
 aaagaattca cccaccagt gcaggctgcc tatcagaag tgggtgctgg tgtggctaact  
 gccttgggcc acaagtatca ctaagctcgc tttcttctgt tccaatttct attaaagggt

cccttggtcc ctaagtccaa ctactaaact gggggatatt atgaagggcc ttgagcatct  
ggattctgcc taataaaaa catttatctt cattgc

SEQ ID NO:14

5 gcatggggac ggcgcttctc cagcgcgggg gctgctttct cctgtgcctt tcgctgctgc  
tcctgggctg ctgggcggag ctgggcagcggg

SEQ ID NO:15

10 cgaaacgatt caaacctct ttactgccgt tatttgctgg atttttattg ctgttttatt  
tggttctggc aggac

SEQ ID NO:16

15 ggagctggg ggaggcttag tgcagcctgg agagtccctg aaactctct gtgaatccaa  
tgaatacgaa ttccctccc atgacatgtc ttgggtccgc aagactccgg agaagaggt  
ggagttggtc gcagccatta atagtgtgg tggtagcacc tactatccag acaccatgga  
gagcgttc atcatctcca gagacaatc caagaagacc ctgtacctgc aaatgagcag  
tctgaggtct gaggacacag ccttgtatta ctgtgcaaga cacagatga gcaaaagtta  
ctgtgagctc aaactaaac ctctgcaga gcattccagg ccagcagggg gcgcggagag  
acacagagtt gtgaaat

SEQ ID NO:17

20 acatccattc ttctgtgagt ttactcgaa gagcagctg tcaactcgga caagccagc  
agctcaccat ggtggacct ccagggtac cagacctctg ggaactggcc ctgagcctca  
cttcacggac acaggctgcc cgccaaagt ggtctcagag caacagtgtg tgcattgctc  
gtcacatctt cctcttgctt tgcagtactg actacacca agaagtgtgc cctggggagg  
aaagcatatt tggcaaccag atcataataa aatcagaaat gcagcaaac tttaaatat  
ccagacttgg

SEQ ID NO:18

30 tggagcaag agggagtatg ctaacttcat g

SEQ ID NO:19

atcaattaca agaggg

SEQ ID NO:20  
atgaagttmg catactccct cttgcttcca ttggcaggag tcagtgcctc agtkatcaat  
tacaagaga

5 SEQ ID NO:21  
aattcttaat taattattgt ggtgtcaca taacttttc

10 SEQ ID NO:22  
ccccccggat ccatggccgc taaattcgtc gtggttcttg ccgcttgctg cgccttgagc  
cactcggcta tggcgccgc caagaagaac ggctac

15 SEQ ID NO:23  
ccccccggat ccatgaaact cctggctcgt ttgcacatgt gcgtgccgc tgccagcgc

20 SEQ ID NO:24  
cagtgtacgg cggctcgagg cagaagtccg gacgcata

25 SEQ ID NO:25  
gaattcatta tcagcacgga ccagatttct ggatcaggat agaagtctga cgtttacggt  
tttcgaagca cagaacgcat tcgttagcgt aagtgtcacc gtcggtacog caaacccggac  
ggtattccag agtgcaaccg ttcagttcgt tgtagcattt agcttcacga ccagagaggt  
ccatggatcc cccttcgct gtcttctcag ttccaagcat tgcgattttg ttaagcaacg  
cactctgat tcgtagagcc tcgttgctgt tgtttgcacg aaccatatg

30 SEQ ID NO:26  
ttcacaggca ggttttttga gagaggggca tgcacatagtc ctcactgtgg

35 SEQ ID NO:27  
aagcttctcg tgaaaaccaa cccaattagt tagtattgca ttctgtgtac tatagtttgg  
aatattaaaa atattttaaa atacctccat ttgtcttacc cttttagtga agatgatacc  
tgcaaaagac atggctaaaag ttatgattgt catgttggga attgttttc ttacaaaac  
ggatgggaaa tctgttaagt aagtactgtt ttgccttggga attggatttt taatgttgac  
tttatcatt cgaagtgggg agctaattgg aagtggccct ctctgtttct ctcttccca  
ggaagagat

SEQ ID NO:28  
atggctacag gctcccgag gtcctgctc ctggctttt gctgctctg cctgccctgg  
cttcaagagg gcagtgc

5

SEQ ID NO:29  
atgaggctct ttgctaatact tggctctttg  
cttcctgcc ctggctgctc tgggg

10

SEQ ID NO:30  
atgc acctgagaat ccacgcgaga cggaacctc ctgcgcggc ggccctggacg cttgggatct  
ggtcccttt ctggggatgt atcgtcagct ct

15

SEQ ID NO:31  
atg gccattagt gagtcctgt gctaggattt ttcatcatag ctgtgtgat gagcgctcag  
gaatcatggg ctatcaaga agaactatgt atcatccagg ccgagttcta tctgaatcct  
gaccaatcag gcgagtttat gtttgac

20

SEQ ID NO:32  
aggggggac ccgcgagacc ttcggttagc aactgtcacc ttgatgctgg cgatcctgag  
ctctcactg gctgagggc

25

SEQ ID NO:33  
atg gtgtgtctga ggctccctgg aggtcctgc atggcagttc tgacagtgac  
actgatggtg ctgagctccc cactggcttt ggct

30

SEQ ID NO:34  
1 gaacgattta aggagcgaat actactggta aactaatgga agaatctgc tgcaccactg  
61 gatattggga gtgtgtgga tgcacctca tcatcaggaa actcaaaaa agaaccgagt  
121 ggtgctagcc aaacagctgt tgttagcga attgttagaa catctctgg agaaggacat  
181 catcaccttg gaaatgagg agctcatcca ggccaaagt ggcagtttca gccagaatgt  
241 ggaactcctc aactgtctgc ctaagagggg tcccaagct ttgatgctc tctgtgaagc  
301 actgagggag accaagcaag gccacctgga ggatatgttg ctaccaccc ttctgggct  
361 tcagcatgta ctcccaccgt tgagctgtga ctacgacttg agtctcctt ttccgggtgtg

421 tgaagtccctgt cccttttaca agaagctccg cctgtcgaca gatactgtgg aacactccct  
481 agacaataaa gatggctcctg tctgcttca ggtgaagcct tgcactcctg aattttatca  
541 aacacacttc cagctggcat ataggttgca gtctcggcct cgtggcctag cactgggtgtt  
601 gagcaatgtg cacttcactg gagagaaaga actggaattt cgctctggag gggatgtgga  
661 ccacagtact ctagtaccc tcttcaagct ttggggctat gactccatg ttctatgtga  
721 ccagactgca caggaaatgc aagagaaact gcagaatttt gcacagttac ctgcacaccg  
781 agtcacggac tctgcatcg tggcactcct ctgcacatggt gtggagggcg ccactctatgg  
841 tgtggatggg aaactgctcc agctccaaga ggtttttcag ctctttgaca acgccaactg  
901 cccaagccta cagaacaaac caaaaatggt cttcatccag gctgcccgtg gagggtgctat  
961 tggatccctt gggcactcc ttctgttcac tgcgtccacc gctctcttg ctctatgaga  
1021 ctgactcgtg ggttgaccaa caagatggaa agaaccacgc aggatccctt ggggtgcgagg  
1081 agagtgatgc cggtaaagaa aagttgcga agatgagact gccacgcgc tcagacatga  
1141 tatcggtcta tgcctgctc aaagggaactg ccgcca.gcg gaacacaaa cgagggttccct  
1201 ggtacatga ggtcttgct caagtgttt ctgagcggc ttgtgatag cactggtccg  
1261 acatgctggt taagtgaaac gcacttatca aggatcggga aggttatgct cctgggcacag  
1321 aattccaccg gtgcaaggag atgtctgaat actgcagcac tctgtgcgc cactcttacc  
1381 tgttccaccg aacccctccc acatgatgtc acctcccat catccacgcc aagtggaaagc  
1441 cactggacca caggaggtgt gatagagcct ttgatcttca ggatgcacgg ttctgttctt  
1501 gcccctcag ggatgtggga atctccaga cttgtttcct gtgcccata tctctgcctt  
1561 tgagtgtggg actccaggcc agtccctttt ctgtgaagcc ctttgcctgt agagccagcc  
1621 ttggttggac ctattgcccag gaatgtttca gctgcagttg aagagcctga caagtgaagt  
1681 tgtaaacaca gtgtggttat ggggagaggg catataaatt ccccatattt ggttcaagt  
1741 ccagcttttg tagatggcac tttagtgatt gcttttatta cattagttaa gatgtctgag  
1801 agaccatctc ctatctttta ttccattcat atccctcgc ctttttgtcc tagagtgaga  
1861 gtttggaaag tgtccaaatt taatgtagac attatctttt ggctctgaag aagcaaacat  
1921 gactagagac gcacttgct gcagtgtcca gaagcggcct gtgcgttccc ttcagtactg  
1981 cagcgcacc cagtggaaag acactcttgg ctcgttttgg ctcaaggcac cgcagcctgt  
2041 cagccaacat tgccttgcat ttgtacctta ttgatctttg cccatggaag tctcaagat  
2101 ctttcgttgg ttgtttctct gagctttgtt actgaaatga gcctcgtggg gagcatcaga  
2161 gaagggcagg aagaatggtg tgtttcccta gactctgtaa ccactctct gtcttttcc  
2221 ttccctgaaa acgtccatct ctctcccta ctattccac ttccattcaa tcaacctgca  
2281 ctccatatct agatttctag aaaagcttcc tagcttatct cctgtcttca tatctctccc  
2341 ttctttacct tcaattcatc ctgttggtg ctgcccacaa atctgtctag aatcctgctt  
2401 tacaggatca tgaatatgct caaagatgta atgtagttct ttgttccctg ttctctttc  
2461 agtattaaac tctcctttga tattatgtgg cttttatttc agtgccatac atgttattgt



2521 ttccaacctt gaaacctttt tccctgctta tctgaaactt cccaacttcc ctgttcttta  
 2581 agactttttt tttttttttt agctcactgc agctcactgc tccgtctgtc gccagggtg  
 2641 gagggaagt gacagatctc agctcactgc agctcactgc tccgggttc acgccattct  
 2701 cctgctcag ccttccaagt agctgggact acagtgccc gccaccgtgc ccggctaatt  
 2761 ttttctgatt tttagtagag acaggggttc accatgttag ccgggatggt ctgactctcc  
 2821 tgactctatg atccaccac ctagcctcc caaagtgtg ggattacagg cgtgagccac  
 2881 tgcccccggg caagaccttt ttttaaaaa aaaaaaaa aaacttccat tcttcttcc  
 2941 tccagtctgt tctacataa cagagtgtt ttggttttt atttttttt gtgttttggc  
 3001 gttttttgtt ttttaagggt agttctact atgtttctc gactgggttc gaactcctgg  
 3061 cctcaagcca ttttcccgc ttagcctctc aaatagctgg gcttacaggc atgagccacc  
 3121 acacctggcc aggatttggg tgtttaata taaatctgat caccctctg cttagaaccc  
 3181 ttctgcttcc tattacctt catttaaaat gtaaaactct cacttgggt tatgagaact  
 3241 ggttcttggc tttcccttga acctaatat atggtgttt cttgctaagc tccagccga  
 3301 gtggtctcct ctcagcttct aatttctgc tcttctctg cctttctctg ggccttctca  
 3361 gctctccacc cccaccctc ttgactcagg tgggtctctt cttcctcaag tcttgacaat  
 3421 tccccggccc ttcagtcctt gacagctcta cttctgtgtc tgtaccaca tcttgtcttt  
 3481 tccccctatt gcatttatt cagtttatat atatgtact tttactgtt catttctgtc  
 3541 tccccctacca ggctgtaaat gaggccagaa acctgtttt tttattcac catcatgtac  
 3601 caagtgtctg gcacatagt ggcttctatt aaatgtttgt tgaataaaag agggaagaag  
 3661 gcaagccaac cttagctaca atctacctt ttgataaaa gtctctttt acaatatata  
 3721 cggattatta tttgtacttt gttttccat gtgttttggc tttatccact ggcattttta  
 3781 gctccttgaa gacatatcat gtgtgagata acttcttca catctccat ggtccttagc  
 3841 aaatgctag gcctgtagta gtcaagggtc tcaataaata tttgtttggg tggtttgtga  
 3901 gccttgctgc caagtctgc ctttgggtcg acatagtatg gaagtatttg agagagagaa  
 3961 ctttccact cccactgcca ggattttgta ttgccatcgg gtgcaaaaata aatgctcata  
 4021 tttttaaaa aaaaaaaaaa aaaa

SEQ ID NO:35

1 tccagataat ctgtctctac caccaggcc atggtgtctt cagcgactat ctgcagtttg  
 61 ctactctca gcagtctctg gatggacatg gccatggcag gtccagctt cttgagccca  
 121 gaccaccaga aagccacaga gaaaaaggaa tccaagaagc caccagtaa actgcagcca  
 181 cgagtcttgg aaggtggct ccaccagag gacagaggac aagcagaaga ggcagaggag  
 241 gagctggaaa tcaggttcaa tgctccctc gatgtggca tcaagctgc aggaagctcag  
 301 taccagcagc atggccgggc cctgggaaa ttctctcagg atctctctg ggaagaggc  
 361 aaagaggcgc cagctaaca gtaaccactg acaggactgg tccctgtact ttcctcttaa

421 gcaagaactc acatccagct tctgcctcct ctgcaactcc cagcactctc ctgctgactt  
481 acaataaat gttcaagctg t

5 SEQ ID NO:36  
atgaagctgc ttgcaatgggt tgcactgctg gtcaccatct gtagcctaga aggagctttg  
gttcggagac

10 SEQ ID NO:37  
atgaagctgc ttgcaatgggt tgcactgctg gtcaccatct gtagcctaga aggagctttg  
gttcggagac

15 SEQ ID NO:38  
atggcctt gccaacggct cgaccctgt tggggtcctg tggagcccc gccctcggca  
gcctcctgtt cctgctcttc agctcggat ggggtgcagcc c

20 SEQ ID NO:39  
atgcccgcgc tgttctccta cctcctagggt gtctggctgc tctgagcca acttcccaga  
gaaatcccag gc

SEQ ID NO:40  
atgacagc atcacttgc gttttaccat cgctttgggt aatattaatt atttttactg  
caccctatac tcactgt

25 SEQ ID NO:41  
atgaaagtc ctgctttgtg acctgtgct gctcagtctc ttctccagtg tgttcagcag  
ttgtcagagg gactgtctca catgccagga gaagctccac ccagccctgg acagttcga  
cctggagggtg tgcatacctcg agtgcgaaga gaaggctctc ccagccccc tctggactcc  
atgcaccaag gtcatggcca ggagctcttg gcagctcagc cctgccccc cagagcatgt  
gggggctgct ctctaccagc cgagagcttc ggagatgcag catctgcggc gaatgccccg  
agtccggagc ttgttccagg agcagggaaga gcccagacct ggcattggagg aggcctgggtga  
gatggagcag aagcagctgc agaagagat

30

SEQ ID NO:42

aattcatgaa gtgggttact ttcatctctt tgttgttctt gttctctctt gcttactcta  
gaggtgtttt cagacg

5

SEQ ID NO:43

atgaagtggg ttactttcat ctctttgttg ttctgtttct ctctgtctta ctctagaggt  
gttttcagac gc

10

SEQ ID NO:44

1 gaattctcaa tggcaaggc aagtgtacat tataaatagc aaacagctg gcttggacca  
61 tgttgccgc cagtcacca gttgagggt ttgaatgaca tcataacct caagagggta  
121 ttgctagcca gctggtgtta ttagaatac acaaaaaca gagaaagaaa acacactctg  
181 gcacacagac tccctctgtc atacacacac acacacacac acacacacac  
241 agaggtttga gttatatgga aaattcaaac aacaggaaaa ttgtttgcc ccaggtacc  
301 ctctctccag agtgggtggg tggggagggg acagtgcag gcagcctagt agaagaataa  
361 agaaaaatgt tctatttcag ttgggtttta cagctcggca tagtcttgc ctcatcgag  
421 gagaaaaagt atgagacagt gccataaagg gaccaatcca atgctgctg cccctccata  
481 cgttcttaga aatgagatca caccctcac ttggcaactg ggacaagggg tcaccgaggt  
541 gctgtcttcc aatctacttt acccagtcac cttcagggtt aaaattgtag agtttgctgg  
601 agagggctct atcgtccttt ctctcttttt ttgtttttaa taatgcattt gctctagaat  
661 ctaaaaattgc tctcccatcc cccatatttc tttaatactg gtaagggtga ttacagacg  
721 ttgtgtctt catgcccagc agaaagttaa tcagaaaaa gatccttatt ttctatggca  
781 gcataagtat tttaatgtct gcgaaccttg tcagtaacac acattctttt aagggaataa  
841 aatgcttctg tgctctagtt ttaaaatgca aaggtatgat gttatttgc accatgccc  
901 aaaaagtct tactcaataa ctttgccaga aggggagag agagagaagg caaatgttcc  
961 ccagctgtt tctgtctac agtgtctgtg tttttagat aaatgtagg attttgtga  
1021 aatccctctt ctgtttgcta aatctcactg tcaactgctaa attcagagca gatagacct  
1081 gcgcaatgga ataaagtctt caaaattgaa atgtgacatt gctctcaaca tctccatct  
1141 ctctggattt ctcttctgtt cattattcct gctaaccaat tcatcttcag actttgtact  
1201 tcagaagcaa tgggaaaaat cagcagctctt ccaacccaat tatcttaagt ctgcttttgt  
1261 gatttcttga aggtaaatat ttcttactct ttgaagtcat tggggaattc ggateccact  
1321 gtaataatag catctttcat ttccgtagta aacgtttcta gatattttgt ctcaattcat  
1381 tgaatatagga acccataaag aaaggggttc agggaggact cctccaaaga tccacagtag  
1441 ccaggggaat aaacacaggt tgttggatgc cgagacacgc tccatccaca actccctgct

1501 ggggttctcat gtactctatt ggcttctgtg ctgggtagtc ctgattaatg acagtctgtgg  
 1561 aatcgtggga gtcaatgcac ttctgtccca cccactccc cttgcaagga tcaaggagga  
 1621 aacctgaacc tcctctgttt tcttgggcag gtgaagatgc acaccatgtc ctctctgcat  
 1681 ctcttctacc tggcgtgtg cctgctcacc ttcaccagct ctgccacggc t

5

SEQ ID NO:45  
 atgaag ccaattcaaa aactcctagc tggccttatt ctactgactt cgtgcgtgga  
 aggtgctcc agccagcact ggtcctatgg actgcgcctt ggaggaaaga gagatgccga  
 aaattt

10

SEQ ID NO:46  
 atgagatt tccttcaatt ttactgcag tttattcgc agcatcctcc gcattagctg  
 ctccagtcac cactacaaca gaagatgaaa cggcacaatt ccggctgaa gctgtcatcg  
 gtactttaga tttagaagg gatttcgatg ttgtgtttt gccattttcc aacagcacia  
 ataacgggtt attgtttata aatactacta ttgccagcat tgctgctaaa gaagaagggg  
 tatctttgga taaaagagag gctgaagctt

15

SEQ ID NO:47  
 atgaa gtgggtaacc tttatttccc ttctttttct ctttagctcg gcttattcca ggggtgtgtt  
 tctctgaga

20

SEQ ID NO:48  
 1 atgaagatgg tctcctctc ggcctccgc tgctcctcgc tgctcctgct gtccctgacc  
 61 gcctccatca gctgctcctt cgcgggacag agagactcca aactcgcct gctgctgcac  
 121 cggtagccgc tgcagggtc caaacaggac atgactcgtt cgccttggc cgagctgtc  
 181 ctgtcggacc tcctgcagg ggagaacgag gctctggagg aggaactt cctctggcc  
 241 gaaggaggac ccgaggagc ccacgccgac cttagagcgg cgcacagcgg ggggcctctg  
 301 ctgcgcctcc gggagagaaa ggcggctgc aagaacttct tctggaaaaa cttcacctcc  
 361 tgctga

25

SEQ ID NO:49  
 atggccgggc gagggggcag cgcgctgctg gctctgtgcg gggcactggc tgcctgcggg  
 tggctcctgg gcgccgaag ccaggagccc ggggcgcctg cggcgggcac gaggcggcgc  
 cggcggctgc agcaagagga cggcatctcc ttcagtagtacc accgtacccc cgagctgcgc

30

gaggcgctcg tgcctgtgtg gctgcagtcg accgccatca gcaggattta cacgggtggg  
 cgcagcttcg agggccggga gctcctggc atcgagctgt ccgacaacc tggcgtccat  
 gagcctggg agcctgaatt taaatacatt gggaaatagc atgggaatga ggctgttga  
 cgagaactgc tcattttctt ggccagtagc ctatgcaacg aataccagaa ggggaacgag  
 acaattgtca acctgatcca cagtaccgcg attcacatca tgccttcctt gaaccagat  
 ggctttgaga aggcagcgtc tcagcctggg gaactcaagg actggtttgt gggtcgaagc  
 aatgcccagg gaatagatct gaaccgggaa tttccagacc tggataggat agtgtacgtg  
 aatgagaaag aaggtggctc aaataatcat ctgttgaata atagaagaa aattgtggat  
 caaaacacaa agcttgctcc tgagaccaag gctgtcattc attggtttat ggatattcct  
 tttgtgcttt ctgccaatct ccatggagg gacctgtgg ccaattatcc atatgatgag  
 acgcggagtg gtagtgtca cgaatacagc tctcccag atgacgccat tttccaaagc  
 ttggcccggt catactctt tttcaaccg gcatgtctg acccaatcg gccaccatgt  
 cgcaagaatg atgatgacag cagcttttga gatggaacca ccaacgggtgg tgcctgggtac  
 agcgtacctg gagggatgca agacttcaat taccttagca gcaactgttt tgagatcacc  
 gtggagctta gctgtgagaa gttcccact ctaccttag gaagagactc tgaagacctc ctgggaggat  
 aacaaaaact cctcattag ctaccttag cagatacacc gagaggttaa aggatttgtc  
 cgagaccttc aaggtaacc aattgcgaat gccaccatct cgttggaaag aatagaccac  
 gatgttacat ccgcaaaagga tgggtgattac tggagattgc ttataacctg aaactataaa  
 cttacagcct cagctccagg ctatctggca ataacaaaga aagtggcagt tcttacagc  
 cctgtgctg ggggtgattt tgaactggag tcattttctg aaaggaaaga agaggagaag  
 gaagaattga tgggaatggg gaaaatgatg tcagaaactt taaatttt

SEQ ID NO:50

atggctctct cactcttcac tgttggaaca ttaattttct tattttggac actcagaatc  
 actgaagcc

SEQ ID NO:51

atgaacaaa ctagcaattc tcgtatcat cgctatggtta ctttcagcg caaacgcctt  
 cagactccaa agcagattga gatcaaatat ggaagcttct gccagag

SEQ ID NO:52

a tggtcagtg gtgcaggctc ttgctgggtg ctgcttgct gctgtgttg caagcacagc  
 tgtctttctc tcagcactgg tctcatggct ggtacctgg aggaagaga gaaatcgact  
 cctacagctc accagagata tctggggaga ttaaactgtg tgaagcggga gaatgcagct  
 atctcaggcc actgaggacc aacatcctaa agagatcctt gattgacacc cttgcaagga

aattccaaaa gaggaatga

SEQ ID NO:53

1 tgtgttttgt agataaatgt gaggatttcc tctaaatccc tcttctgctt gctaaatctc  
61 actgtgcgtg ctaaatccag agcagataga gctgcgcaa tcgaataaa gtctcaaaa  
121 ttgaatgtg actttgctct aacatctccc atctctctgg attctctttt gctcattat  
181 tctgcccac caattcattt ccagactttg tacttcagaa gcatgggga aaatcagcag  
241 tcttccaact caattattta agatctgctt ctgtgacttc ttgaagataa agatacacat  
301 catg

10

SEQ ID NO:54

atgtc aggcccgagg acgtgcttct gtctaccgtc ggctcttgta ctagtactgc tgaagtctcag  
cacttcggca ctaggg

15

SEQ ID NO:55

mspaqlaka aarstcmtrl psgirvatap snshfaavgv yvdagpiyet sidrgvshfv  
sslafksthg atesqvlktm aglgnlfct atresilyqg svlhhdiprt vqlladttlr  
palteeiae rratiafeae dlhsrpdafi gemmhavafg grglgnsifc epqrarnmts  
dtireyfaty lhpsrmvav tgvahaelvd lvskafvpss trapssvths dietayvgs  
hqlvipkppp thpnyeqtlc hvqvafvpvp fthpdmfpvs tlqvlmgggg afsagggpgkg  
mysrlytnvl nryrwmesca afqhayssts lfgisascvp sfnpnlcnvl agefvhmarn  
lsdeevarak nqlkssllmn lesqvltved igrqvlaqng rleplelvnn isavtrd'lv  
rvaealvakp ptmvavgedl tkltidiketl aafnasgeal qpvgsgsagfig rvtm

20

25

SEQ ID NO:56 (pTnMCS)

1 ctgacggcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga  
61 ccgctacact tgcagcgcc ctacgcccgc ctccttctgc tttctccct tctttctcg  
121 ccacgtctgc cggcatcaga ttggctattg gccattgcat acgttgatc catatcataa  
181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac  
241 tagttattaa tagtaatacaa ttacggggtc attagttcat agccatata tggagttccg  
301 cgttacataa cttacggtaa atggcccgc ttcccatagt aacgccaata gggactttcc attgacgtca  
361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca  
421 atgggtggag tattttacggg aaactgcccc ctgggcagta catcaagtgt atcatatgcc  
481 aagtagcccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta  
541 catgacctta tgggactttc ctacttgga gtacatctac gtattagtca tcgctattac  
601 catggtgatg cgggttttggc agtacatcaa tggcggtgga tagcgggttg actcacgggg

30

35

661 atttccaagt ctccacccca ttgaogtcaa tgggagtttg ttttggcacc aaatcaacg  
 721 ggactttcca aaatgtcgta caactcgc ccaatggcg caaatggcg gtaggcgtgt  
 781 acggtgggag gtctatataa gcagagctcg ttagtgaac cgtcagatcg cctggagacg  
 841 ccatccaagc tgttttgacc tccatagaag acacgggac gcatccagcc tccgcggccg  
 901 ggaacggtgc atggaacgc ggaattcccg tgccaagagt gacgtaagta ccgcctatag  
 961 actctatagg cacacccctt tggctcttat gcattctata ctgttttttg cttggggcct  
 1021 atacaccccc gtttccttat gctataggtg atggtatagc ttacctata ggttggggtt  
 1081 attgaccatt attgaccact ccoctattgg tgacgatact ttccattact aatccataac  
 1141 atggctcttt gccacaacta tctctattgg ctatatgcca atactctgc cttcagagac  
 1201 tgacacggag tctgtatttt tacaggatgg tttccattt attattaca aattcacata  
 1261 tacaacaacg ccgtccccc tgccgcagc tttattaaa catagcgttg gatctccacg  
 1321 cgaatctcgg gtacgtgttc cggacatggg ctctctccg gtagcggcg agctccaca  
 1381 tccgagccct ggtcccatgc ctccagcggc tcatggtgc tggcagctc cttgctccta  
 1441 acagtggagg ccagacttag gcacagca atgcccacca rcaccagtg gccgcacaag  
 1501 gccgtggcgg tagggatagt gtctgaaat gagcgtggag attgggctcg cacggctgac  
 1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgtgtattc  
 1621 tgataagagt cagaggtaac tcccggtgag gtgctgttaa cggctggagg cagtgtagtc  
 1681 tgagcagtag cgtgtgctgc gcgcgcgc accagacata atagctgaca gactaacaga  
 1741 ctgttctctt ccatgggtct tttctcagt caccgtgga coactgtgga actcgatatt  
 1801 ttacacgact ctctttacca attctgccc gaattacact taaaacgact caacagctta  
 1861 acgttggtt gccacgcat acttgactgt aaactctca ctcttacoga acttggccgt  
 1921 aacttgcca ccaagcgag acaaaacat aacatcaaac gaatcgaccg attgttaggt  
 1981 aatcgtcacc tccacaaga gcgactcgt gtataccgtt ggcagctag ctttatctgt  
 2041 tgggggaata cgaatgccat tgaacttgtt gactggtctg atattcgtga gcaaaaaaga  
 2101 ctatgtgtat tggagcttc agtcgacta cagggtcgtt cgtttactct ttatgagaaa  
 2161 gcgttcccg ttcagagca atgttcaag aaagctcatg accaatttct agccgacctt  
 2221 gcgagcattc taccagata aatccgttga gaactgggt tggactggt taagtctagt aagaggaaaa  
 2281 gtacaatatg cagaccatgg agcggaaaac tggaaaccta tcagcaactt acatgatag  
 2341 tcatctagtc actcaaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca  
 2401 tgccaaattc tattgtataa atctcgtctt aaaggccgaa aaatcagcg ctcgacacgg  
 2461 actcatgttc accaccgtc acctaaaac tactcagct cggcaaaagg gccatgggtt  
 2521 tagcaacta acttacctgt tgaatttoga acacccaaac aacttgttaa tatctattcg  
 2581 aagcgaatgc agatgaaga aaccttcga gacttgaata gtctcgcta cggactaggc  
 2641 ctacgcata gccgaacgag cagctcagag cgttttgata tcatgtctgt aatcgccctg  
 2701 atgcttcaac taacatgttg gcttgcggc gttcactgtc agaacaagg ttgggacaag  
 2761 cacttccagg ctaacacagt cagaaatcga acgtactct caacagttcg cttaggcatg  
 2821 gaagttttgc ggcattcttg ctacacaata acaagggaag acttactcgt ggtgcaacc  
 2881 ctactagctc aaatttatt cacacatggt tacgctttgg ggaattatg aggggatcgc

5

10

15

20

25

30

35

3001 tctagagcga tccgggatct cgggaaaagc gttggtgacc aaaggtgcct tttatcatca  
 3061 ctttaaaaat aaaaacaat tactcagtcg ctgttataag cagcaattaa ttatgattga  
 3121 tgcctacatc acaacaanaa ctgatttaac aaatggttgg tctgacctag aaagtatat  
 3181 tgaacattat ctgtattata ttattgataa taataaaaac ctatcccta tccaagaagt  
 3241 gatgcctatc attggttggg atgaacttga aaaaaattag ccttgaatac attactggta  
 3301 aggtaaaagc catgttcagc aaattgatcc aagagaacca acttaagct ttcctgacgg  
 3361 aatgttaatt ctctgttgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaatc  
 3421 attaaagttaa ggtggatata catcttgtca tatgatcccc gtaatgtgag ttacttcact  
 3481 cattaaggcac ccaggctttt acattttatg ctccggctc gtatgttggg tggaaattgtg  
 3541 aggggataac aatttcacac agaaaacagc tatgacctag attacgcaa gcgcgaatt  
 3601 aacctcact aaagggaaac aaagctggag ctccacggcg gtggcgcccg ctctagaact  
 3661 agtggatccc cggggtgca ggaattcgat atcaagctta tggataccgc tgacctcgag  
 3721 gggggggcccg gtacccaatt cgcctatag tgagtcgtat tacgcggct cactggccgt  
 3781 cgttttaca cgtcgtgact gggaaaaccc tggcgttacc raacttaac gccttgacg  
 3841 acatccccc ttcgccagct ggcgtaatag cgaagaggcc cgcaccgac gcccttccc  
 3901 acagttggcg agcctgaatg gcgaatggaa attgtlaagcg ttaatatattt gttaaaaatc  
 3961 gcgtaaat tttgttaaat cagctcattt tttaaaccaat aggcgaaat cggcaaaatc  
 4021 ccttataaat caaagaata gaccgagata gggttgagtg tttgtccagt ttggaacaag  
 4081 agtccactat taagaacgt ggactccaac gtcaagggc gaaaaaccgt ctacagggc  
 4141 gatggccac tactcggga tcatatgaca agatgtgtat ccaccttaac ttaatgattt  
 4201 ttacaaaaat cattagggga ttcatcagtg ctacgggtca acgagaatta acattccgtc  
 4261 aggaagcctt atgatgatga tttgtctaaa aacttactca atggctgggt atgcatatcg  
 4321 caatacatgc gaaaaccta aaagagcttg ccgataaaaa aggccaat tttgtatatt  
 4381 acccgggctt tttattgagc ttgaaagata aataaaatag atagggtttta ttgaaagcta  
 4441 aatcttcttt atcgtaaaaa atgcctctt gggttatcaa gagggctcatt atatttcgcg  
 4501 gaataaacatc atttgggtgac gaaataacta agcacttgtc tctgtttac tccctgagc  
 4561 ttgaggggtt aacatgaagg tcatcgatag caggataata atacagtaaa acgctaaacc  
 4621 aataatccaa atccagccat cccaaattgg tagtgaatga ttataataa cagcaaacag  
 4681 taatggggca ataacaccgg ttgcatgtgt aaggctcacc aataatcccc gtaaaagcac  
 4741 ttgtgtatga ctctttgttt ggatagacat cactccctgt aatgcaggta aagcgatccc  
 4801 accaccagcc aataaaatta aaacaggga aactaaccaa ccttcagata taacgctaa  
 4861 aaaggcaaat gcactactat ctgcaataaa tccgagcagt actgcccgtt ttcgcccc  
 4921 ttatgtgcta tcttctctgc cacaaaggct tggaaatactg agtgtaaaaag accaagcccc  
 4981 taagtgaana gccaccatc atgctatca gcgctgaaat aataatcaac aaatggcgc gttaaataag  
 5041 gtgtgtgatt ggctatcaat cgtgtgaaat ttgttccctt tagtgaggggt taattgagcg  
 5101 ttatgtatatac cgatcagctt ttgttccctt ttgttccctt ttgttccctt ttgttccctt  
 5161 tcatggtcat agctgtttcc ttgttccctt ttgttccctt ttgttccctt ttgttccctt  
 5221 cgagccggaa gcaataaagt taaagcctgg ggtgcctaat ggtgagccta actcacatta  
 5281 attcgtgttc gctcactgcc cgctttccag tcgggaaacc tgcgtgtgcca gctgcaataa

5

10

15

20

25

30

35



5341 tgaatcgcc aacgcgcgg gagaggcggg ttgcgtattg ggcgtcttc cgcttctctg  
 5401 ctcactgact cgctgcgtc ggtcgttcgg ctcgcgcgag cgttatcagc tcactcaaa  
 5461 gcggtaatac ggttatccac agaatacagg gataacgcag gaaagaacat gtgagcaaaa  
 5521 gcccagcaaa aggcacagaa ccgtaaaag ccgcgttgc tggcgtttt ccataggctc  
 5581 cgccccctg acgacatca caaaaaaga cgtcaagtc agagtggcg aaaccgaca  
 5641 ggactataa gataccagg gttccccct ggaagctccc tcgtgcgtc tcctgttccg  
 5701 accctgcgc ttaccggata cctgtccgc ttctccctt cgggaagcgt ggcgtttct  
 5761 catagctcac gctgtaggta tctcagttcg gcccagccg tgcgcttat cgttcttgag  
 5821 gtgcacgaac cccccgttca gcccagccg tgcgcttat cgttcttgag  
 5881 tcaaccccg taagacacga cttatcgcca ctggcagcag ccactggtaa cagattagc  
 5941 agagcagggt atgtaggcgg tgcacagag ttcttgaagt ggtggcctaa ctacggctac  
 6001 actagaagg cagtatttgg tatctgcgt ctgctgaagc cagttacett cggaaaaaga  
 6061 gttggtagct cttgatccg caaacaacc accgtggta ggggtgggtt tttgtttgc  
 6121 aagcagcaga ttacgcgcg aaaaaagga tctcaagaag tctctttgat ctttcttca  
 6181 ggtctgcag ctcagtggaa cgaaaactca cgttaaggga ttttggctat gagattatca  
 6241 aaaaggatct tcactagat ccttttaaat taaaaatgaa gttttaaata atctaaagt  
 6301 atatatgagt aaacttggtc tgacagttac caatgcttaa tcagtggagg acctatctca  
 6361 gcgactgtc tatctgttc atccatagtt gctgaactcc cgtctgtga gataactacg  
 6421 atacggcagg gcttaccatc tggccccagt gctgcaatga taccgcgaga cccacgcitca  
 6481 cgggtccag atttatcagc ataaaccag ccagccggaa gggccgagcg cagaagtgg  
 6541 cctgcaact tatccgcctc catccagctc attaatgtt gccgggaagc tagagtaagt  
 6601 agtcgcagc ttaatagttt gcgcaacgtt gttgccattg ctacagcat cgtgggttca  
 6661 cgtcgtcgtc ttggatggc ttcattcagc tccggttccc aacgatcaag gcgagttaca  
 6721 tgatccccc tgtgtgcaaaa aaagcgggt agtctcttcg gtcctccgat cgttgtcaga  
 6781 agtaagtgg ccgcagtggt atcaactatg gttatggcag cactgcataa ttctcttact  
 6841 gtcattgcat ccgtaagatg ctttctctg actggtgagt actcaacca gtcattctga  
 6901 gaatagtga tgcggcgacc gagtgtctc tgcgcggcgt caatacggga taataccgcg  
 6961 ccacatagca gaactttaa agtgcctatc attgaaaaa gttctcggg gcgaaaaact  
 7021 tcaaggatct taccgtgtt gagatccagt tcgatgtaac ccactcgtc acccaactga  
 7081 tcttcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg aaggcaaaat  
 7141 gccgcaaaa agggataaag ggcgacacgg aaatgttgaa tactcatact cttctctttt  
 7201 caatattatt gaagcattta tcagggttat tgtctcatga gcggatacat atttgaatgt  
 7261 atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt gccac

SEQ ID NO:57 (pTnMod)

CTGACGCGCC CTGTAGCGGC GCATTAAAGC CGCGGGGTGT GGTGGTTACG 50

CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTGCG 100  
TTTCTTCCTT TCCTTTCTCG CCACGTTGCG CGGCATCAGA TTGGCTATTG 150  
GCCATTGCAT ACGTTGTATC CATATCATAA TATGTACATT TATATTGGCT 200  
CATGTCACAA ATTACCGCCA TGTTGACATT GATTATTGAC TAGTTATTAA 250  
TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG 300  
CGTTACATAA CTTACGGTAA ATGGCCCGCC TGSCGTGACCG CCCAACGACC 350  
CCCGCCCATTT GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA 400  
GGGACTTTCC ATTACGTCA ATGGGTGGAG TATTACGGT AAAGTGCACA 450  
CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCC CCTATTGACG 500  
TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCAGTA CATGACCTTA 550  
TGGGACTTTT CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC 600  
CATGGTGATG CGGTTTGGC AGTACATCAA TGGGGTGGG TAGCGTTTG 650  
ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG 700  
TTTTGGCACC AAAATCAACG GGAATTTCCA AATGTGTA ACAACTCCG 750  
CCCATTTGAG CAAATGGCG GTAGGCGTGT ACGTGGGAG GTCTATATAA 800  
GCAGAGCTCG TTTAGTGAAC CGTCAGATCG CCTGGAGACG CCATCCACGC 850  
TGTTTTGACC TCCATAGAAG ACACCGGGAC CGATCCAGCC TCCGCGGCCG 900  
GGAACGGTGC ATTGGAACGC GGAATCCCCG TGCCAAAGAT GACGTAAGTA 950  
CCGCCCTAAG ACTCTATAGG CACACCCCTT TGGCTCTTAT GCATGCTATA 1000  
CTGTTTTTGG CTTGGGGCCT ATACACCCCG GCTTCCTTAT GCTATAGGTG 1050  
ATGGTATAGC TTAGCCTATA GGTGTGGGTT ATTGACCATT ATTGACCACT 1100  
CCCCATTGG TGACGATACT TTCCATTACT AATCCATAAC ATGGCTCTTT 1150  
GCCACAATA TCTCTATTGG CTATATGCCA ATACTCTGTC CTTCAGAGAC 1200  
TGACACGGAC TCTGTATTTT TACAGGATGG GGTCCCATTT ATATTTTACA 1250  
AATTCACATA TACAACAACG CCGTCCCCCG TGCCCCGAGT TTTTATTAAA 1300  
CATAGCGTGG GATCTCCACG CGAATCTCG GTACGTGTC CGGACATGGG 1350  
CTCTTCTCCG GTAGCGGGCG AGCTTCCACA TCCGAGCCCT GGTCCCATGC 1400  
CTCCAGCGGC TCATGGTCG TCGGCAGCTC CTTGCTCCTA ACAGTGGAG 1450  
CCAGACTTAG GCACAGCACA ATGCCACCA CCACAGTGT GCCGCACAAG 1500  
GCCGTGGCG TAGGTATGT GTCTGAAAAT GAGCGTGGAG ATTGGGCTCG 1550  
CACGGCTGAC GCAGATGGAA GACTTAAGGC AGCGGCAGAA GAAGATGCAG 1600  
GCAGCTGAGT TGTGTATTTC TGATAAGAGT CAGAGGTAAC TCCCGTTGCG 1650  
GTGCTGTTAA CGGTGGAGGG CAGTGTAGTC TGAGCAGTAC TCGTTGCTGC 1700  
CGCGCGCGCC ACCAGACATA ATAGCTGACA GACTAACAGA CTGTTCTTTT 1750  
CCATGGGTCT TTTCTGCAGT CACCGTCGGA CCATGTGTGA ACTTGATATT 1800

5 TTACATGATT CTCTTTACCA ATTCTGCCCC GAATTACACT TAAAACGACT 1850  
 CAACAGCTTA ACGTTGGCTT GCCACGCATT ACITGACTGT AAAACTCTCA 1900  
 CTCCTACCGA ACTTGGCCGT AACCTGCCAA CCAAAGCGAG AACAAAACAT 1950  
 AACATCAAC GAATCGACCG ATTGTAGGT AATCGTCACC TCCACAAAGA 2000  
 GCGACTCGCT GTATACCGTT GGCATGCTAG CTTTATCTGT TCGGGAATAC 2050  
 GATGCCCAAT GTACTTGTG ACTGGTCTGA TATTCGTGAG CAAAAACGAC 2100  
 TTATGGTATT GCGAGCTTCA GTCGCACTAC ACGTCTGTC TGTACTCTT 2150  
 TATGAGAAAG CGTTCCCGCT TTCAGAGCAA TGTTCAAAGA AGCTCATGA 2200  
 CCAATTCTA GCCGACCTTG CGAGCATCT ACCGAGTAAC ACCACACCG 2250  
 10 TCATTGTGAG TGATGCTGGC TTAAAGTGC CATGGTATAA ATCCGTTGAG 2300  
 AAGCTGGGTT GGTACTGGTT AAGTCGAGTA AGAGGAAAAG TACAATATGC 2350  
 AGACCTAGGA GCGGAAAACCT GGAACCTAT CAGCAACTTA CATGATATGT 2400  
 CATCTAGTCA CTCAAAGACT TTAGGCTATA AGAGGCTGAC TAAAAGCAAT 2450  
 CCAATCTCAT GCCAAATTCT ATTGTATAA TCTCGCTCTA AAGCCGAAA 2500  
 15 AAATCAGGCG TCGACACGGA CTCATTGTCA CCACCCGTCA CCTAAAATCT 2550  
 ACTCAGGCTC GGCAAAGGAG CCATGGGTTT ACTTGTAAAT ATCTATTCGA 2600  
 GAAATTCGAA CACCCAAACA ACCTTCCGAG ACTTGAAAAG TCCTGCCTAC 2650  
 GATTGAAGAA ACCTTCCGAG CCGAACGAGC AGCTCAGAGC GTTTTGATAT 2700  
 TACGCCATAG CCGAACGAGC TGCTTCAACT AACATGTTGG CTTCGGGGC 2750  
 20 ATCGCCCTGA TGCTTCAACT AACATGTTGG CTTCGGGGC TTAACATCGA 2800  
 GAAACAAGGT TGGGACAAGC ACTTCCAGGC TTAGGCATGG AAGTTTTCG 2850  
 ACGTACTCTC AACAGTTCCG TTAGGCATGG GCTGCAACCC TACTAGCTCA 2900  
 TACACAATAA CAAGGGAAGA CTTACTCGTG GAAATTATGA TAATGATCCA 2950  
 25 AAATTATTC ACACATGGTT ACGCTTTGGG TCTAGAGATC TGTGTGTTGG 3000  
 GATCACTTCT GGCTAATAAA AGATCAGAGC CTTTCTAGTT GCCAGCCATC 3050  
 TTTTCTGTGG ATCTGCTGTG TGCCTTCCCTT GACCTTGGAA GGTGCCACTC 3100  
 CCTTCCCCCG TGCTTCTCTT TCGCATCGCA TTGTCTAGT AGGTGTCATT 3150  
 TTCTTAATAA AATGAGGAAA TTAGCATCGCA GCAAGGGGGA GGATGGGAA 3200  
 30 CTATTCTGGG GGTGGGGTG GGGCAGCACA GCATCTATGG GTACCTCTCT 3250  
 GACAATAGCA GGCATGCTGG GGATCGGGTG GGTCTCTCTCT CTCTCTCTCT 3300  
 CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CCGTACCAGG TGCTGAAGAA 3350  
 CTCTCTCTCT CTCTCTCTCT GCCTTTTATC ATCATTCTAA AATAAAAAA 3400  
 TTGACCGGGT GACCAAGGT GATGCTGTTA TAAGCAGCAA TTAATTATGA 3450  
 CAATTACTCA GTGCTGTTA TAACTGATT TAACAAATGG TTGGTCTGCC 3500  
 35 CATCACAACA AAAACTGATT TTAGAAAGTA TTAGAAAGTA 3550

5 TATTGGAACA TTATCTTGAT TATATTATTG ATAATAATAA AAACCTTATC 3600  
CCTATCCAAG AAGTGATGCC TATCATTTGGT TGAATGAAC TTGAAAAAAA 3650  
TTAGCCTTGA ATACATTACT GGTAAGGTAA ACGCCATTGT CAGCAAATTG 3700  
ATCCAAGAGA ACCAACTTAA AGCTTTCTTG ACGGAATGTT AATTCTCGTT 3750  
GACCCTGAGC ACTGATGAAT CCCCTAATGA TTTTGGTAAA AATCATTAAAG 3800  
TTAAGGTGGA TACACATCTT GTCATATGAT CCCGGTAATG TGAGTTAGCT 3850  
CACTCATTAG GCACCCCCAGG CTTTACACTT TAATGCTTCCG GCTCGTATGT 3900  
TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC 3950  
CATGATTACG CCAAGCGCGC AATTAAACCT CACTAAAAGG AACAAAAGCT 4000  
GGAGCTCCAC CGCGTGGCG GCGCTCTAG AACTAGTGA TCCCCGGGC 4050  
TGCAGGAAAT CGATATCAAG CTTATCGATA CCGCTGACCT CGAGGGGGG 4100  
CCCGGTACCC AATTGCCCC ATAGTGAGTC GTATTACGCG CGTCACTGG 4150  
CCGTCGTTTT ACAACGTCGT GACTGGGAAA ACCCTGGCGT TACCCAACTT 4200  
AATCGCCTTG CAGCACATCC CCCTTTCGCC AGCTGGCGTA ATAGCGAAGA 4250  
GGCCCGCACC GATCGCCCTT CCCAACAGTT GCGCAGCCTG AATGGCGAAT 4300  
GGAAATTGTA AGCGTTAATA TTTTGTAAAA ATTCCGTTA AATTTTGT 4350  
AAATCAGCTC ATTTTAAAC CAATAGGCCG AAATCGGCAA AATCCCTTAT 4400  
AAATCAAAAG AATAGACCGA GATAGGGTTG AGTGTGTTT CAGTTTGGAA 4450  
CAAGAGTCCA CTATTAAAGA ACGTGGACTC CAACGTCAA GGGCGAAAA 4500  
CCGTCATCA GGGCGATGCG CCACTACTCC GGGATCATAT GACAAGATGT 4550  
GTATCCACCT TAACCTTAATG ATTTTACC AATCATTAG GGGATTCTATC 4600  
AGTGCTCAGG GTCAACGAGA ATTAACATTC CGTCAGGAAA GCTTATGATG 4650  
ATGATGTGCT TAAAAACTTA CTCAATGGCT GGTTATGCT ATCGCAATAC 4700  
ATGCGAAAA CCTAAAAGAG CTTGCCGATA AAAAAGGCCA ATTTATTGCT 4750  
ATTTACCGCG GCTTTTATTT GAGCTTGAAA GATAAATAAA ATAGATAGGT 4800  
TTTATTGAA GCTAAATCTT CTTTATCGTA AAAAATGCC TCCTGGGTTA 4850  
TCAAGAGGGT CATTATATTT CGCGGAATA CATCATTTGG TGACGAAATA 4900  
ACTAAGCACT TGTCTCCTGT TTACTCCCTT GAGCTTGAGG GGTAAACATG 4950  
AAGTCAATCG ATAGCAGGAT AATAATACAG TAAAACGCTA AACCAATAAT 5000  
CCAAATCCAG CCATCCCAA TTGGTAGTGA ATGATTATAA ATAACAGCAA 5050  
ACAGTAATGG GCCAATAACA CCGGTGCTAT TGGTAAAGGT CACCAATAAT 5100  
CCCTGTAAG CACCTTGCTG ATGACTCTTT GTTTGGATAG ACATCACTCC 5150  
CTGTAATGCA GGTAAGCGA TCCCACCACC AGCCAATAAA ATTAACACAG 5200  
GGAAAACTAA CCAACCTTCA GATATAAACG CTAAAAAGGC AAATGCACTA 5250  
CTATCTGCAA TAAATCCGAG CAGTACTGCC GTTTTTCGC CCATTAGTG 5300

5 GCTATTCTTC CTGCCACAAA GGCTTGGAAAT ACTGAGTGTA AAAGACCAAG 5350  
ACCCGTAATG AAAAGCCAAAC CATCATGCTA TTTCATCATCA CGATTTCTGT 5400  
AATAGCACCA CACCGTGCTG GATTGGCTAT CAATGGCGTG AAATAATAAT 5450  
CAACAAATGG CATCGTTAAA TAAGTGATGT ATACCGATCA GCTTTTGTTT 5500  
CCTTTAGTGA GGGTTAATTG CGCGCTTGGC GTAATCATGG TCATAGCTGT 5550  
TTCTGTGTG AAATGTATTAT CCGCTCACAA TTCCACACAA CATAAGGCC 5600  
GGAAGCATAA AGTGTAAGC CTGGGGTGCC TAATGAGTGA GCTAACTCAC 5650  
ATTAATTGGG TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA AACCTGTGCT 5700  
GCCAGCTGCA TTAATGAATC GGCCAACGCG CGGGAGAGG CGGTTTGGCT 5750  
ATTGGGGCGT CTTCGCTTC CAGCTCACTA AAGGCGGTA ATACGGTTAT 5800  
TCGGCTGCGG CGAGCGGTAT CAGCTCACTC AATGTGAGC AAAAGGCCAG 5850  
CCACAGATC AGGGGATAAC GCAGGAAGA ACATGTGAGC AAAAGGCCAG 5900  
CAAAAGGCCA GGAACCGTAA AAAGGCCGG TTGCTGGCGT .TTTCCATAG 5950  
GCTCCGCCCC CCTGACGAGC ATCACAAAAA TCGACGCTCA AGTCAGAGGT 6000  
GGCGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC CCTGGGAAGC 6050  
TCCCTCGTGC GCTCTCCTGT TCCGACCCCTG CCGCTTACCG GATACCTGTC 6100  
CGCCTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCATAGC TCACGCTGTA 6150  
GGTATCTCAG TTCGCTGTAG GTCGTTGCTT CCAAGCTGGG CTGTGTGCAC 6200  
GAACCCCGG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA ACTATCGTCT 6250  
TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG 6300  
GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTCTTG 6350  
AAGTGGTGGC CTAACCTACGG CTACACTAGA AGGACAGTAT TTGGTATCTG 6400  
CGCTCTGCTG AAGCCAGTTA CCTTCGGRAA AAGAGTTGGT AGCTCTTGAT 6450  
CCGGCAACA AACCAACGCT GGTAGCGGTG GTTTTTTTGT TTGCAAGCAG 6500  
CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCCTT TGATCTTTTC 6550  
TAGGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA GGGATTTTGG 6600  
TCATGAGATT ATCAAAAAGG ATCTTCACT AGATCCCTTTT AAATTAATAA 6650  
TGAAGTTTAA ATCAATCTA AAGTATATAT GAGTAAACTT GGTCGTGACAG 6700  
TTACCAATGC TTAATCAGTG AGGCACTAT CTCAGCGATC TGTCTATTTC 6750  
GTTTCATCCAT AGTTGCCTGA CTCCCGTCTG TGTAGATAAC TACGATACCG 6800  
GAGGGCTTAC CATCTGGCCC CAGTGTGCA ATGATACCGC GAGACCCACG 6850  
CTCACCGGCT CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG 6900  
AGCGCAGAAG TGGTCTGCA ACTTTATCCG CCTCCATCCA GTCTATTAAAT 6950  
TGTTGCCGGG AAGCTAGAT AAGTAGTTCT CCAGTTAATA GTTTGCGCAA 7000  
CGTTGTTGCC ATTGCTACAG GCATCGTGGT GTCACGCTCG TCGTTTGGTA 7050

TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGCGAGT TACATGATCC 7100  
 CCCATGTTGT GCAAAAAGC GGTAGCTCC TTGGGTCCCTC CGATCGTTGT 7150  
 CAGAAGTAAG TTGGCCGCAG TGTATCACT CATGGTTATG GCAGCACTGC 7200  
 ATRATTCTCT TACTGTCAAG CCATCCGTAA GATGCTTTTC TGTGACTGGT 7250  
 5 GAGTACTCAA CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTTG 7300  
 CTCTTGCCCG GCGTCAATAC GGGATAATAC CGCGCCACAT AGCAGAACTT 7350  
 TAAAAGTGCT CATCATTTGA AAACGTTCTT CGGGGCGAAA ACTCTCAAGG 7400  
 ATCTTACCGC TGTTGAGATC CAGTTCGATG TAACCCACTC GTGCACCCAA 7450  
 CTGATCTTCA GCATCTTTTA CTTTCACCCAG CGTTTCTGGG TGAGCAAAA 7500  
 10 CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAATGT 7550  
 TGAATACTCA TACTCTTCCT TTTTCAATAT TATTGAAGCA TTTATCAGGG 7600  
 TTATTGTCTC ATGAGCGGAT ACATATTTGA ATGTATTTAG AAAAATAAAC 7650  
 AAATAGGGGT TCCGCGCACA TTTCCCGGAA AAGTGCCAC 7689

15 SEQ ID NO:58 (a Kozak sequence)  
 ACCATGG

SEQ ID NO:59 (a Kozak sequence)  
 ACCATGT

20 SEQ ID NO:60 (a Kozak sequence)  
 AAGATGT

25 SEQ ID NO:61 (a Kozak sequence)  
 ACGATGA

SEQ ID NO:62 (a Kozak sequence)  
 AAGATGG

30 SEQ ID NO:63 (a Kozak sequence)  
 GACATGA

TGAATGTGTT CTGTGTTAT CAATATAAT CACAGTTAGT GATGAAGTTG GCTGCAAGCC  
 TGCATCAGTT CAGCTACTTG GCTGCATTTT GTATTGTTT CTGTAGGAAA TGCAAAAAGGT  
 CTTAGGCTGA CCTGCATTG TATCCTCTTT GCCTTACTGC TGAGAATCTC TGCAGGTTTT  
 AATATGTTTAC ATTTCGTCC CATTACTTTT GGAAGATAAA ATATTACAG AATGCTTATG

AAACCTTTGT TCATTTTAAAA ATATTCCTGG TCAGCGTGAC CGAGCTGAA AGAACACATT  
 GATCCCGTGA TTTCATPAAA TACATATGTT CCATATATTG TTTCTCAGTA GCCTCTTAAA  
 TCATGTGCGT TGGTGACAT ATGAATACAT GAATAGCAAA GGTTCATCTG GATTACGCTC  
 TGGCCTGCAG GAATGGCCAT AAACCAAGC TGAGGGAAGA GGGAGAGTAT AGTCAATGTA  
 GATTATACATG ATTGCTGATT GGGTTATTAT CAGCTAGATA ACAACTTGGG TCAGGTGCCA  
 GGTCAACATA ACCTGGGCAA AACCACTC ACATCAATTT TAAATTTATT GTAAATGCCG  
 GCCGTGACCC AATCTAGGAA AGCAAGTAGC AACTTCTGGT CAATCAGAAA AAGTTTTTTT  
 TAGTAGAAGT GTTTTACTGT GATACATTGA AACTTCTGGT CAATCAGAAA AAGTTTTTTT  
 ATCAGAGATG CCAAGGTATT ATTTGATTTT CTTTATTCGC CGTGAAGAGA ATTTATGATT  
 GCAAAAAGAG GAGTGTTTAC ATAAACTGAT AAAAAACTTG AGGAAATTCAG CAGAAAAACAG  
 CCACGTGTTT CTGAACATTC TTCCATAAAA GTCTCACCAT GCCTGGCAGA GCCCTATTCA  
 CCTTCGGT

SEQ ID NO:70 (fragment of ovalbumin promoter - chicken)

GAGGTCAGAAAT GGTTCCTTTA CTGTTTGTC AATCTATTAT TTCAATACAG  
 AACAAATAGCT TCTATAACTG AAATATAATTT GCTATTGTAT ATTATGATTG  
 TCCCTCGAAC CATGAACACT CCTCCAGCTG AATTCACAA TTCCTCTGTC  
 ATCTGCCAGG CCAITTAAGTT ATTCAATGGA GATCTTTGAG GAACACTGCA  
 AGTTCATATC ATAAACACAT TTGAAATGA GTATTGTTT GCATGTATG  
 GAGCTATGTT TTGCTGTATC CTCAGAAAAA AAGTTTGTTA TAAAGCATTC  
 ACACCCATAA AAAGATAGAT TTAATATTC CAGCTATAGG AAAGAAAGTG  
 CGTCTGCTCT TCACTCTAGT CTCAGTTGGC TCCTTCACAT GCATGCTTCT  
 TTATTTCTCC TATTTGTCA AGAAAAATAT AGGTCACGTC TTGTTCTCAC  
 TTATGTCCTG CCTAGCATGG CTGAGATGCA CGTTGTAGAT ACAAGAAGGA  
 TCAAAATGAAA CAGACTTCTG GTCTGTTACT ACAACCATAG TAATAAGCAC  
 ACTAACTAAT AATTGCTAAT TATGTTTTCC ATCTCTAAGG TTCCCACATT  
 TTTCTGTTTT CTTAAGATC CCATTATCTG GTTGTAACAG AAGTCAATG  
 GAACATGAGC AATATTTCCC AGTCTTCTCT CCCATCCAC AGTCTGATG  
 GATTAGCAGA ACAGGCAGAA AACACATTGT TACCCAGAAT TAAAAACTAA  
 TATTGCTCT CCAATTCATC CAAATGGAC CTATTGAAAC TAAATCTAA  
 CCAATCCCA TTAATGATT TCTATGGCGT CAAAGGTCAA ACTTCTGAAG  
 GGAACCTGTG GGTGGGTCAC AATTCAGGCT ATATATTCCC CAGGCTCAG

SEQ ID NO:71 (chicken ovalbumin enhancer)

ccgggctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag



cttgaccta tacttgattt tcttcaaaact ggggaaacaa cacaatccca caaaacagct  
 cagagagaaa ccatcactga tggctacagc accaaggtat gcaatggcaa tccattcgac  
 attcatctgt gacctgagca aaatgattta tctctccatg aatggttgct tctttccctc  
 atgaaaaggc aatttccaca ctcaaatat gcaacaaaga caaacagaga acaattaatg  
 tgctccttcc taatgtcaaa attgtagtgg caaagaggag acaaaaatct caagttctga  
 gtaggtttta gtgattggat aagaggcttt gacctgtgag ctacacctga cttcatatcc  
 ttttggataa aaagtgcctt tataactttc aggtctcoga gtctttattc atgagactgt  
 tgggttaggg acagaccac aatgaaatgc ctggcatagg aaagggcagc agagccttag  
 ctgacctttt cttgggacaa gcaattgtcaa acaatgtgtg acaaaaactat ttgtactgct  
 ttgcacagct gtgctgggca gggcaatcca ttgccacctc tcccagggtaa ctttccaact  
 gcaagaagat tgttgcttac tctctctaga

5

10

SEQ ID NO:72 (5' untranslated region)

GTGGATCAACATACAGCTAGAAAGCTGTATTGCCCTTTAGCACTCAAGCTCAAAAGACAACCTCAGAGTTTCAC  
 C

15

8

SEQ ID NO:73 (putative cap site)

ACATACAGCTAG AAAGCTGTAT TGCCTTTAGC ACTCAAGCTC AAAAGACAAC TCAGAGTTCA

20

SEQ ID NO:74 (Chicken Ovalbumin Signal Sequence)

ATG GGCCTCATCG GCGCAGCAAG CATGGAATTT TGTITTGATG TATTCAAGGA GCTCAAAAGTC  
 CACCATGCCA ATGAGAAACAT CTTCTACTGC CCCATTGCCA TCATGTCAGC TCTAGCCATG  
 GTATACCTGG GTGCAAAAGA CAGCACCAGG ACACAGATAA ATAAGGTTGT TCGCTTTGAT  
 AAACCTCCAG GATTCCGGAGA CAGTATTGAA GCTCAGTGTG GCACATCTGT AAACGTTTAC  
 TCTTCACTTA GAGACATCCT CAACCAAATC ACCAAACCAA ATGATGTTTA TTCGTTTCAGC  
 CTTGCCAGTA GACITTTATGC TGAAGAGAGA TACCCAATCC TGCCAGAATA CTTGCAGTGT  
 GTGAAGGAAC TGTATAGAGG AGGCTTGGAA CCTATCAACT TTCAAACAGC TGCAGATCAA  
 GCCAGAGAGC TCATCAATTC CTGGGTAGAA AGTCAGACAA ATGGAATTAT CAGAAATGTC  
 CTTCAGCCAA GCTCCGTGGA TTCTCAAACT GCAATGGTTC TGGTTAATGC CATTGTCTTC  
 AAAGGACTGT GGGAGAAAAC ATTTAAGGAT GAAGACACAC AAGCAATGCC TTTTCAGAGTG  
 ACTGAGCAAG AAAGCAAAAC TGTGCAGATG ATGTACCAGA TTGGTTTATT TAGAGTGGCA  
 TCAATGGCTT CTGAGAAAAT GAAGATCCTG GAGCTTCCAT TTGCCAGTGG GACAATGAGC  
 ATGTTGGTGC TGTTCCTTGA TGAAGTCTCA GGCCTTGAGC AGCTTGAGAG TATAATCAAC  
 TTTGAAAAAC TGACTGAATG GACCAGTTCT AATGTTATGG AAGAGAGGAA GATCAAAAGTG  
 TACTTACCTC GCATGAAGAT GGAGGAAAAA TACAACCTCA CATCTGTCTT AATGGCTATG

25

30

35

GGCATTACTG ACGTGTGTTAG CTCTTCAGCC AATCTGTCTG GCATCTCTC AGCAGAGAGC  
 CTGAAGATAT CTCAAGCTGT CCATGCAGCA CATGCAGAAA TCAATGAAGC AGGCAGAGAG  
 GTGGTAGGCT CAGCAGAGGC TGGAGTGGAT GCTGCAAGCG TCTCTGAAGA ATTTAGGGCT  
 GACCATCCAT TCCTCTTCTG TATCAAGCAC ATCGCAACCA ACGCCGTTCT CTTCITTGGC  
 AGATGTGTTT CCCCT

5

SEQ ID NO:75 (Chicken Ovalbumin signal sequence-shortened approx. 50bp)  
 ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG TATCAAG

10 SEQ ID NO:76 (Chicken Ovalbumin signal sequence-shortened approx.  
 100bp)

ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG TATCAAGGA GCTCAAAAGTC  
 CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA TC

15 SEQ ID NO:77 (vitellogenin targeting sequence)

ATGAGGGGATCATACTGGCATTAGTGCTCACCCCTGTAGGCAGCCAGAGTTTGACATTGGT

SEQ ID NO:78 (p146 protein)

KYKKALKKLAKLL

20

SEQ ID NO:79 (p146 coding sequence)

AAATACAAAAAGCACTGAAAAAACTGGCAAAACTGCTG

SEQ ID NO:80 (pro-insulin sequence)

TTTGTGAACCAACACCTGTGGGCTCACACCTGGTGAAGCTCTCTACCTAGTGTGCGGGACGAGGCTT  
 CTTCTACACACCCAGACCCGCCGGAGGAGCAGAGACCTGCAGGTGGGCGAGGTGGAGCTGGGCGGGGGCC  
 CTGTTGCAGGCAGCCTGCAGCCCTTGGCCCTGGAGGGGTCCCTGCAGAACGCTGGCATTTGTGGAAACAATGC  
 TGTACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTCTGCAACTAG

25

30 SEQ ID NO:81 (spacer)  
 (GPGG)<sub>x</sub>

SEQ ID NO:82 (spacer)  
 GPGGPGGPGPGG

35

SEQ ID NO:83 (spacer)  
GGGGGGGGGGGGG

5 SEQ ID NO:84 (spacer)  
GGGGGGGGGGGGGGG

SEQ ID NO:85 (repeat domain in TAG spacer sequence)  
Pro Ala Asp Asp Ala

10 SEQ ID NO:86 (TAG spacer sequence)  
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala  
Pro Ala Asp Asp Ala Pro Ala Asp Asp

15 SEQ ID NO:87 (gp41 epitope)  
Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu

SEQ ID NO:88 (polynucleotide sequence encoding gp41 epitope)  
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly Ser  
Cys Gly Trp Ile Gly Leu Leu Asp Asp Asp Lys

20 SEQ ID NO:89 (enterokinase cleavage site)  
DDDDK

SEQ ID NO:90 (TAG sequence)  
25 Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala  
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp  
Ile Gly Leu Leu Asp Asp Asp Lys

30 SEQ ID NO:91 (altered transposase Hef forward primer)  
ATCTCGAGACCATGTGTGAACCTTGATATTTACATGATTCTCTTACC

SEQ ID NO:92 (altered transposase Her reverse primer)  
GATTGATCATATCATATAATTCCCAAGCGTAACC

5 SEQ ID NO:93 (Xho I restriction site)  
CTCGAG

SEQ ID NO:94 (Bcl I restriction site)  
TGATCA

10  
SEQ ID NO:95 (CMVf-NgoM IV primer)  
TTGCCGGCATCAGATTGGCTAT

15  
SEQ ID NO:96 (Syn-polyAr-BstE II primer)  
AGAGGTCACGGGTCAATTCTTCAGCACCTGGTA

102  
SEQ ID NO:97 (pTnMOD (CMV-prepro-HCPro-Lys-CPA))  
1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga  
61 ccgtacacct tgcacgcgcc cttagcggcg ctcctttcgc tttcttccct tcccttctcg  
121 ccaggttcgc cggcatcaga ttggctattg gccattgcat acgttgatc catatcataa  
181 tatgtacatt tatattggct catgtccaac attaccgcca tgtgacatt gattattgac  
241 tagttattaa tagtaataa ttacggggtc attagttaat agcccatata tggagtccg  
301 cgttacataa cttacggtaa atggccggcc tggctgacgc cccaacgacc ccggtccatt  
361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca  
421 atgggtggag tatttacggt aaactgcca cttggcagta catcaagtgt atcatatgcc  
481 aagtagcggc cctattgacg tcaatgacgg taaatggccc gctggcatt atgccagta  
541 catgacctta tgggactttc ctacttgga gtacatctac gtattagtca tcgtattac  
601 catggtgatg cggttttggc agtatatcaa tgggcgtgga tagcgggttg actcacgggg  
661 atttccaagt ctccaccca ttgacgtcaa tgggagtttg tttggcacc aaatcaacg  
721 ggaatttcca aaatgcgta acaactcgc ccattgacg caaatgggcy gtagcgtgt  
781 acgttgggag gtctatataa gcagagtcg tttagtgaac cgtcagatcg cctggagacg  
841 ccattccacgc tgttttgacc tccatagaag acaccggggac cgtaccagcc tccgcgggcg  
901 ggaacgggtgc attggaacgc ggattccccc tgcacaagat gacgtaagta ccgcctatag  
961 actctatagg cacacccctt tggctcttat gaatgctata ctgttttttg cttggggccct  
1021 atacaccccc gcttccctat gctatagggt aggttatagc ttagcctata ggtgtgggtt  
1081 attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac  
1141 atggtctctt gccacaacta tctctattgg ctatatgcca atactctgc cttcagagac

1201 tgacacggac tctgtatttt tacaggatgg ggtccattt attattaca aattcacata  
1261 tacaacaacg cgtcccccg tgcgcgagt ttattataa catagcgtgg gatctccacg  
1321 cgaatctcgg tactgtgttc cggacatggg tctctcccg gtagcggcg agcttcacac  
1381 tccgagccct ggtcccatgc ctcacagcgc tcctgtcgc tcggcagctc cttgtccta  
1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag  
1501 gccgtggcgg tagggtagt gtctgaaat gagcgtggag attgggctcg cacggctgac  
1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgtgtattc  
1621 tgataagagt cagaggtaac tccgttgcg gtctgttaa cggcggaggg cagtgtagtc  
1681 tgagcagtag cgtttgctgc cgcgcgcgc accagacata atagctgaca gactaacaga  
1741 ctgttccctt ccattggctt ttctgcagt cacgctcggg ccatgtgtga acttgatatt  
1801 ttacatgatt ctctttacca attctgccc gaattacact taaacgact caacagctta  
1861 acgttggctt gccacgcatt acttgactgt aaaaacttca ctctaccga acttggccgt  
1921 aacctgcca ccaagcgag acaaaacat aacatcaac gaatcgacc attgttaggt  
1981 aatcgtcacc tccacaaga gcgactcgt gtatacrgt ggcgtgctag ctttatctgt  
2041 tcgggcaata cgtgcccatt tgtacttgt gactgtctg atattcgtga gcaaaacga  
2101 cttatgggat tgcgagcttc agtcgcacta cacggtcgtt ctgttactct ttatgagaaa  
2161 gcgttcccgc ttccagagca atgttcaag aaagctcatg accaatttct agccgacctt  
2221 gcgagcattc taccagtaga caccacccg ctcatgtca gtgatgctgg ctttaaatg  
2281 ccatgggata aatcgttga gaagctgggt tggactcgtt taagtcgagt aagaggaaa  
2341 gtacaatatg cagacctagg agcgaaaaac tggaaacctt cagcaactt acatgatatg  
2401 tcatctagtc actcaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca  
2461 tgcaaatc tattgtataa atctcgtctt aaaggccgaa aaaatcagcg ctgcacacgg  
2521 actcattgtc accaccgtc acctaaatc tactcagcgt cggcaaaagg gccatgggtt  
2581 ctagcaacta acttacctgt tgaattcga acacccaaac aacttgttaa tatctatcog  
2641 aagcgaatgc agattgaaga aaccttccga gacttgaana gtccctgcta cggacaggc  
2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgt atcgccctg  
2761 atgcttcaac taacatgtt gcttgcgggc gtctatgctc agaaacaagg ttgggacaag  
2821 cacttccagg ctacacagat cagaatcga acgtactct caacagttcg cttaggcattg  
2881 gaagttttgc ggcattctgg ctacacaata caaagggaag acttactcgt ggtgcaacc  
2941 ctactagctc aaattttatt cacacatggt tacgctttgg ggaatttatg ataattatcc  
3001 agatcacttc tggtaataa aagatcagag ctctagagat ctgtgtgttg gtttttttg  
3061 gatctgctgt gccttctagt tgcagccat ctgttgtttg cccctcccc gtgccttct  
3121 tgaccttgga aggtgccact cccactgtcc ttctctaata aaatgaggaa attgcatcgc  
3181 atgtctgag taggtgtcat tctattctgg ggggtgggtt ggggcagcac agcaaggggg  
3241 aggattggga agacaatagc aggcattgct gggatgcggg ggtctctatg ggtacctctc  
3301 tctctctctc tctctctctc tctctctc tctctctc cctctctc tctctctc  
3361 tctctctc tctctctc tctctctc tctctctc tctctctc tctctctc  
3421 tgccttttat catcacttta aaataaaaa caattactc agtgcctgtt ataagcagca  
3481 attaatatg atgatgctt acatcaaac aaaactgat ttaacaatg gttggtctgc

5

10

15

20

25

30

35

3541 cttagaagat atatttgaaac attatcttga ttatatatt gataataata aaaaccttat  
 3601 ccctatccaa gaagtgatgc ctatcattgg ttggaatgaa cttgaaaaaa attagccttg  
 3661 aatacattac tggtaaggta aagccatttg tcagcaaat gatccaagag aaccaactta  
 3721 aagctttcct gacggaatgt taattctcgt tgacocctgag cactgatgaa tcccctaatg  
 3781 atttttgtaa aaatcattaa gttaagggtg atacacatct tgtcatatga tcccggtaaat  
 3841 gtgagttagc tcactcatga ggcaccccgag gcttacact ttatgcttcc ggctcgtatg  
 3901 ttgtgtggaa ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac  
 3961 gccaaaggcg caattaacc tcaataaagg gaacaaaagc tggagctcca cgcgggtggc  
 4021 ggccgctcta gaactagtgg atccccggg ctgcaggaat tcgatatcaa gcttatcgat  
 4081 accgctgacc tggagcatca gattggctat tggccattgc atacgttgta tccatatcat  
 4141 aatatgtaca ttatatattg ctcattgtcca acattaccgc catgttgaca ttgattattg  
 4201 actagtatt aatagtaac aattacgggg tcattagtcc atagccata tatggagtcc  
 4261 cgcgttacat aacttaagg aaatggcccg cctggctgac cgcccaacga ccccgccca  
 4321 ttgacgtcaa taatgacgta tgtcccata gtaacgcaa taggacttt ccattgacgt  
 4381 caatgggtgg agtatttacg gtaaacctgc cacttggcag tacatcaagt gtatcatatg  
 4441 tcaagtacgc cccctattga cgtcaatgac ggtaaatggc cgcctggga ttatgcccag  
 4501 tacatgacct tatgggactt tctacttgg cagtacatct acgtattagt catcgtatt  
 4561 accatgggtga tgcgggtttg gcagtacatc aatgggctg gatagcgggt tgaactcacgg  
 4621 ggatttccaa gtcttccacc cattgacgtc aatgggagtt tgttttgga ccaaaatcaa  
 4681 cgggaacttc caaaagtgc taacaactcc gcccatatga cgcaaatggc cggtaggcgt  
 4741 gtacgggtgg aggtctatat aagcagagct cgtttagtga accgtcagat cgcctggaga  
 4801 cgccatccac gctgttttga cctccataga agacaccggg accgatccag cctccgggc  
 4861 cgggaacggt gcattggaac ggggattccc cgtgccaaga gtgacgtaag taccgctat  
 4921 agactctata ggcacacccc ttbggtcttt atgcatgcta tactgttttt ggcttgggg  
 4981 ctatacacc cgccttcttt atgctatagg tgatgggtata gcttagccta taggtgtggg  
 5041 ttattgacca ttattgacca ctcccctatt ggtgacgata cttccatta ctaatccata  
 5101 acatggctct ttgccacaac tatctctatt ggtatatatgc caatactctg tccctcagag  
 5161 actgacacgg actctgtatt ttacaggat ggggtcccat ttattattta caaatccaca  
 5221 tatacaaaa cgccttcccc cgtgcccga gttttatta aacatagcgt gggatctcca  
 5281 cgcgaatctc gggtaogtgt tccggacatg ggtctcttc cgttagcggc ggagcttcca  
 5341 catccagacc ctggtcccat gctccacgct gctcatggtc gctggcagc tccctgctcc  
 5401 taacagtggg gcccagactt aggcacagca caatgcccac caccaccagt gtgcggcaca  
 5461 aggcgtggc ggtagggtat gtgtctgaaa atgagcgtgg agattgggct cgacggctg  
 5521 acgagatgg aagacttaag gcagcggcag aagaagatgc aggcagctga gttgtgtat  
 5581 tctgataaga gtcagaggta actcccgctg cgtgtctgtt aacggtggag ggcagtgtag  
 5641 tctgagcagt actcgttctt gccgcgcgcg ccaccagaca taatagctga cagactaaca  
 5701 gactgttctt tcccatgggt cttttctgca gtcaccgtcg gatcaatcat tcatctcgtg  
 5761 actctctcgt gtgtgggtgt tacctatata tctaaattta atattctgtt tattaataat  
 5821 taatatattt cgacgatgaa ttctctcaagg atatttttct tctgtctcgc ttggttctctg

5

10

15

20

25

30

35

5881 gctttgtcaa cagtttggc tgccgagag ccgaaggta ccaggtgca gctgcaggag  
 5941 tcggggggag gcttggtaaa gccggggggg tcccttagag tctcgtgag agcctctgga  
 6001 ttcactttca gaaacgctg gatgagctgg gtccgcccag ctccaggaa ggggctggag  
 6061 tgggtcggcc gtattaaaag caaaattgat tggggacaa cagactatgc tgcacccgtg  
 6121 aaaggagat tcaccatctc agagatgat tcaaaaaa cgttatatct gcaatgaat  
 6181 agcctgaaag ccgaggacac agcgtatat tactgtacca cgggattat gataacattt  
 6241 gggggagtta tccctcccc gaattggggc cagggaacc ttgtaccgt ctctcagcc  
 6301 tccaccaagg gcccatcgtt ctctccccg gacccctct ccaagagcac ctctggggg  
 6361 acagcggccc tgggctgctt ggtcaaggac tactccccg aaccggtgac ggtgtcgtg  
 6421 aactcaggc ccttgaccag cggcgtgcac acctttccg ctgtctaca gtctcagga  
 6481 ctctacttcc ttaggcaagt ggtgacctg cctccagca gcttggggc ccagacctac  
 6541 atctgcaacg tgaatcaca gccacgaac accaagggtg acaagaaagt tggcccaaa  
 6601 tctgtgaca aaactcacac atgccaccg tgcacgac ctgaactctt ggggggaccg  
 6661 tcatgttcc tcttcccc aaacccaag gacacctca tgatctccg gacctgag  
 6721 gtcacatgc tgggtgtgga cgtgagccac gaagaacctg aggtcaagt caactggtac  
 6781 gtggacggcg tggaggtgca taatgccaag acaaacccg gggaggagca gtacaacagc  
 6841 acgtaccgtg tggtcagcgt cctcacogtc ctgcaccag actggctgaa tggcaaggag  
 6901 tcaagtga aggtctcaa caaagcctc ccagcccca tgcagaaac catctccaa  
 6961 gccaaaggc agcccgaga accacgggtg tacacctgc cccatccc ggtgagctg  
 7021 accaagaacc aggtcagctt gacctgctg gtcaaggct tctatccc cgacatgcc  
 7081 gtggagtgg agagcaatgg gcagcggag acaactaca agaccacgc tccgtgctg  
 7141 gactccgac gctccttctt cctctacag agctcaccg tggacaagag caggtggcag  
 7201 caggggaacg tcttctcatg ctccgtgatg catgaggtc tgcacaacca ctacacgag  
 7261 aagggcctct cctgtctcc gggtaaacg ccagagccga aaagcttct ctatgaactg  
 7321 acacagccac cctcgtgtc agtgtccca ggaacacgg ccaggtcac ctgct.tgga  
 7381 gatgcatgc cagaaaaata tgtttattg taccagaga agtcaggcca ggcctctg  
 7441 gtggtcatct atgaggacag caaacgacc tccgggatcc ctgagagatt ctctggctcc  
 7501 agctcaggga caatggccac cttgactatc agtggggccc aggtggaaga tgaagggtgac  
 7561 tactactgtt actcaactga cagcagtggt tatcataggg aggtgttcag cggaggggacc  
 7621 aagctgaccg tcttaggtca gcccaaggct gcccctcgg tcaactgtt cccacctcc  
 7681 tctgaggagc tcaagccaa caaggccca ctgggtgtgc tcaatagta ctctaccg  
 7741 ggaagccgtga cagtggcctg gaaggcagat agcagcccc tcaaggcggg agtggagacc  
 7801 accacacct ccaacaaaag caacacaag tacgcggcca gcagctacct gagcctgacg  
 7861 ctgagcagt ggaagtccca caaagctac agctgccag tcaagcatga agggagcacc  
 7921 tgggagaaga cagtggcccc tgcagaatgt tcacgcgga gggagggaag ggccttttt  
 7981 gaagggggag gaaacttgc gccatgactc ctctgtgccc cccgcacgg aacactgatg  
 8041 gtcagagggc cctctgccat tgcgtcttcc tctgccttc ctctgactc tgaatgtggc  
 8101 tctttgtgta ctgcacagc aagaataaa atctcaacat ctcaatgggt tctctgagat  
 8161 ttttcaagag tggtaagca catctctcc ccagaccctc ttgctgagg ccagtgccag

5

10

15

20

25

30

35

8221 gcaccaactt ggctactgct gccatgaga gaaatccagt tcaatatattt ccaagcaaaa  
 8281 atggattaca tatgcctag atoctgatta acaggtgttt tgtattatct gtgctttcgc  
 8341 ttaccacaca ttatccatt gctcccttc gactcgaggg ggggccgggt acccaattcg  
 8401 ccctatagtg agtcgtatta cgcgcgtca ctggccgtcg ttttacaacg tcgtgactgg  
 8461 gaaaaccctg gcgtttacca acttaatcgc cttgcagcac atccccctt cgcagctgg  
 8521 cgtaatagcg aagaggcccg caccgatcgc cttcccaac agttgcgcag cctgaatggc  
 8581 gaatggaat tgaagcgtt aatatattgt taaaattcgc gttaaatttt tgttaaatca  
 8641 gctcattttt taaccaatag gcgaaatcg gcaaaatccc ttataatca aaagaataga  
 8701 ccgagatagg gtgagtggtt gttccagttt ggaacaagag tccactatta aagaacgtgg  
 8761 actcaactg caaaggcgga aaacccgtct atcaggcgga tggccacta ctccgggac  
 8821 atatgacaag atgtgtatcc accttaactt atgattttt accaaatca ttaggggatt  
 8881 catcagtgct cagggtcaac gagaattaac attcgtcag gaaagcttat gatgatgatg  
 8941 tgcctaaaaa cttactcaat ggctggttat gcataatcga ataatcgga aaacctaata  
 9001 agagcttgcc gataaaaaa gcaaatattt tgctatttac cgcggctttt tattgagctt  
 9061 gaaagataaa taaatagat aggtttttat tgaagctaaa tctttttat cgtaaaaaat  
 9121 gccctcttgg gttatcaaga gggctcattt atttcggga ataactcat ttggtgacga  
 9181 aataactaag cacttgctc ctgtttactc cctgagctt gaggggttaa catgaaggtc  
 9241 atcgatagca gataataat acagtaaac gctaaaccaa taatccaat ccagccatcc  
 9301 caaattggta gtgaatgatt ataaataca gcaaacagta atgggccaat aacacgggtt  
 9361 gatttggtaa ggctaccaa taatccctgt aagcacctt gctgatgact ctttgtttgg  
 9421 atagacatca ctccctgtaa tgcaggtaaa ggcgtccac caccagcaa taaattaaa  
 9481 acagggaata ctaaccaacc ttcatatata aacgctaaa aggcataatgc actactatct  
 9541 gcaataaatc cagcagtag tgcgtttttt tgcacctt agtggctatt ctctctgcca  
 9601 caagggtctg gaatactgag tgaataagac caagaccgt aatgaaaagc caacca:cat  
 9661 gctattctat atcagattt ctgtaatagc accacacgt gctggattgg ctatcaatgc  
 9721 gctgaaataa taatcaaca atggcatcgt taaataagtg atgtataccg atcagctttt  
 9781 gttcccttta gtgagggtta attgcgctt tggcgtaatc atggtcatag ctgtttcctg  
 9841 tgtgaaattg ttatcgcctc acaattccac acaacatacg agccggaagc ataaagtga  
 9901 aagcctgggg tgcctaata gtgagctaac tcacattaat tgggttgccg tcactgcccg  
 9961 ctttccagtc gggaaacctg tcgtgcccag tgcattaatg aatcgcccaa cgcgcgggga  
 10021 gaggggggtt gcgtattggg cgctcttccg ctctctcgt cactgactcg ctgectcgg  
 10081 tcgttcggct gcgcgagcg gtatcagctc atccaaaggc ggtaatcag ttatccacag  
 10141 aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc  
 10201 gtaaaaggc cgcgttgctg gcgtttttcc ataggctcgc cccccctgac gacatcaca  
 10261 aaaaatcgac ctcaagtcag aggtgggaa acccgacagg actataaga taccaggcgt  
 10321 ttccccctgg aagctccctc gtgcgtctc ctgttccgac cctgcccgtt accgataacc  
 10381 tgcgcgctt tctcccttcg ggaagcgtgg cgttttctca tagctcagc tgtagggtatc  
 10441 tcagttcggg taggtcgtt cgtcccaagc tgggctggt gcacgaaccc cccgttcagc  
 10501 ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc caaccggga agacacgact

5

10

15

20

25

30

35



10561 tatcgccact ggcagcagcc actgggtaaca ggattagcag agcgagggtat gtaggcgggtg  
 10621 ctacagagtt cttgaagtgg tggcctaact agggctacac tagaaggaca gtatttggta  
 10681 tctggctctt gctgaagcca gttaccttcg gaaaaagagt tggtagctct tgatccggca  
 10741 aacaaaccac cgctggtagc ggtgggtttt ttgtttgcaa gcagcagatt acggcgagaa  
 10801 aaaaaggatc tcaagagat cctttgatct ttctacaggg gtctgacgtc cagtggaaacg  
 10861 aaaaactcac ttaaggatt ttgggtcatga gattatcaaa aggatcttc acctagatcc  
 10921 ttttaaatca aaaaatgaagt tttaaatcaa tctaaagtat atatgagtaa acttgggtctg  
 10981 acagttacca atgcttaatc agtgaggcac ctatctcagc gatctgtcta ttctgttcat  
 11041 ccatagttgc ctgactcccc gtctgtgtaga taactacgat acgggagggc ttaccatctg  
 11101 gccacagtgc tgcaatgata ccgagagacc cagctcacc ggctccagat ttatcagcaa  
 11161 taaaccagcc agcgggaagg gccgagcgca gaagtgttcc tgcaacttta tcggctcca  
 11221 tccagtctat taattgttgc cgggaagcta gagtaagtag ttgcgcagtt aatagtgtgc  
 11281 gcaacgttgt tgccattgct acaggcatcg tgggtgcacg ctgcgtcttt ggtatggctt  
 11341 cattcagctc cggttcccaa cgatcaaggc gagttaratg atccccatg ttgtgcaaaa  
 11401 aagcgggttag ctcttctggt cctccgacgc ttgtcgaag taagttggcc gcagtgttat  
 11461 cactcatggt tatggcagca ctgcataatt ctcttactgt catgccatcc gtaagatgct  
 11521 tttctgtgac tgggtgagtag tcaaccaagt cattctgaga atagtgtatg cggcgaccga  
 11581 ttgtctcttg ccggcgctca atacgggata ataccgccc acatagcaga actttaaaag  
 11641 tgtctatcat tggaaaacgt tcttcggggc gaaaactctc agggatctta ccgctgttga  
 11701 gatccagttc gatgtaacc actctgtcac ccaactgacg ttacgcatct tttactttca  
 11761 cagcagtttc tgggtgagca aaacagggaa ggcaaaatgc cgcaaaaag ggaataaagg  
 11821 gacacgggaa atgttgaata ctcatctct tccttttca atattattga agcattttatc  
 11881 aggggttattg tctcatgagc ggatacatat ttgaatgtat ttagaanaat aaacaaatag  
 11941 ggggtccgcg cacattttcc cgaagaagtgc cac

SEQ ID NO: 98 (pTMCS (CHOVep-prepro-HCPro-CPA))  
 1 ctgacgcgcc ctgtagcgc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga  
 61 ccgctacact tggcagcgc ctaggcgcgc ctctttcgc tcttctccct tctttctctg  
 121 ccacgttctg cggcatcaga ttgggtattg gccattgcat acgttgtatc catatcataa  
 181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac  
 241 tagttattaa tagtaatcaa ttacgggggtc attagtctat agccatata tggagttccg  
 301 cgttacataa cttacggtaa atggccgcgc ttgctgacgc ccaacgacc cccgccatt  
 361 gacgtcaata atgacgtatg ttcccatagt aagccaata gggactttcc attgacgtca  
 421 atgggtggag tatttaccgt aaactgccc cttggcagta cateaagtgt atcatatgcc  
 481 aagtagcccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta  
 541 catgacctta tgggactttc ctacttgcca gtacatctac gtattagtca tgcctattac  
 601 catgggtgat cgggttttggc agtacaatca tagcgggtttg tagcgggtttg actcacgggg  
 661 atttcaagt ctccaccca ttgacgtcaa tgggagtttg ttttggcacc aaatcaaacg  
 721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatggggc gtaggcgtgt

781 acggtgggag gtctatataa gcagagctcg tttagtgaa cgtcagatcg cctggagagc cctggagagc  
 841 ccaccacgc tgtttgacc tccatagaag acaccgggac cgatccagcc cgatccagcc tccgcgccg  
 901 ggaacggtgc attggaacgc ggaattccccc tgccaaagagt gacgtaagta cgcctatag  
 961 actctatagg cacaccctt tggctcttat gcatgctata ctgtttttgg cttggggcct  
 1021 atacacccc gcttccttat gctatagggt atgggtatag ttagcctata ggtgtgggtt  
 1081 attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac  
 1141 atgggtcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac  
 1201 tgacacggac tctgtatttt tacaggatgg ggtcccattt attatttaca aattcacata  
 1261 tacaacaacg ccgtccccc tgcccgcagt ttttattaaa catagcgtgg gatctccacg  
 1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctcgg gtacggcgg agctccaca  
 1381 tccgagccct ggtcccatgc ctccagcggc tcatggtcgc tggcagctc cttgctccta  
 1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag  
 1501 gcgtggcgg tagggtatgt gtctgaaat gagcgtggag attgggctcg cacggctgac  
 1561 gcagatggaa gacttaaggc agcggcagaa gaaga\*gcag gcagctgagt tgttgtattc  
 1621 tgataagagt cagaggtaac tcccggtgcy ggcgtgttaa cgttggaggg cagtgtagtc  
 1681 tgagcagtag tctgtgctgc cgcgcggcc accagacata atagctgaca gactaacaga  
 1741 ctgttcctt ccatgggtct tttctgcagt caccgtcggc ccatgtgcga actcgatat  
 1801 ttacacgact ctctttacca attctgccc gaattacact taaacgact caacagctta  
 1861 acgttggctt gccacgcat acttgactgt aaaacttca ctcttaccga acttggccgt  
 1921 aactggcaa ccaagcgag acaaaaacat aacatcaaac gaatcgacg attgttaggt  
 1981 aatgctacc tccacaaga gcgactcgt gtatacgtt ggcagctag ctttatctgt  
 2041 tggggcaata cgatgcccat tgtacttgt gactggtctg atattcgtga gcaaaaacga  
 2101 cttatgggtat tgcgagcttc agtcgacta caggtcgtt ctgttactct ttatgagaaa  
 2161 gcgttcccc ttccagagca atgttcaaa gaaagctcatg accaatttct agcggacctt  
 2221 gcgagcattc taccagtaa caccacacg ctcatgttca gtgatgctgg ctttaaaagt  
 2281 ccacgggtata aatcgttga gaagctgggt tggactcgtt taagtcgagt aagaggaaa  
 2341 gtacaatatg cagacctagg agcggaaaac tggaaacctc tccgcaactt acatgatatg  
 2401 tcatctagtc actcaaaagac tttaggtat aagaggctga ctaaaagcaa tccaatctca  
 2461 tgcacaattc tattgtataa atctcgtctt aaaggccgaa aaaatcagcg ctccacacgg  
 2521 actcattgtc accaccgtc acctaaaac tactcagcgt cggcaaaagg gccatgggtt  
 2581 ctgcaacta acttacctgt tgaattcga acaccacac aacttgttaa tatctattcg  
 2641 aagcgaatgc agattgaaga aacctccga gacttgaata gtctgccta cggactaggc  
 2701 atcgccata gccgaacgag cagctcagag cgttttgata tcatgtgctt aatcgccctg  
 2761 atgtttcaac taacatgttg gcttggggc gttcatgctc agaacaagg ttgggacaag  
 2821 cacttccagg ctacacagt cagaaatcga aacgtactct caacagttcg cttaggcagt  
 2881 gaagttttgc ggcattcttg ctacacaata acaagggaag acttactcgt ggcgtcaacc  
 2941 ctactagctc aaatttatt cacacatggt tacgctttgg ggaattatg aggggatcgc  
 3001 tctagagcga tccgggatct cgggaaaagc gttgggtgacc aaaggtgcct ttatcatca  
 3061 ctttaaaat aaaaacaat tactcagtc cgtttataag cagcaattaa ttatgattga

5

10

15

20

25

30

35

3121 tgcctacatc acaacaaaaa ctgatttaac aaatggttgg tctgccttag aaagtatat  
 3181 tgaacattat cttgattata ttattgataa taataaaaaa cttatcccta tccaagaagt  
 3241 gatgcctatc attggttggg atgaacttga aaaaaattag ccttgaatac attactggta  
 3301 aggtaaacgc catgtgcgc aaatgatcc aagagaacca acttaaaagt ttcctgacgg  
 3361 aatgtttaatt ctggttgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaatc  
 3421 attaaagttaa ggtggatata catcttgtca tatgatcccc gtaatgtgag ttagctcact  
 3481 cattaggcac cccaggcttt acactttatg cttccggctc gttatgttgg tggaaattgtg  
 3541 agcgataaac aatttcacac aggaacacgc tatgacctg attacgcaa gcgcgaatt  
 3601 aacctcact aaaggaaca aaagtggag ctccacgcg gggcgggcgc ctctagaact  
 3661 agtggatccc cggggctgca ggaattcgat atcaagctta tcatccgcg tgacctcgag  
 3721 ctgcagaaaa atgcaggtg gactatgaac tccatccaa aggagcttga cctgatacct  
 3781 gatttcttc aacagggga acaacacaa tcccacaaa cagctcagag agaaccatc  
 3841 actgatggtc acagcaccaa ggtatgcaat ggaatccat tgcacatcca tctgtgacct  
 3901 gagcaaaatg atttctctc ccatgaatgg ttgct.cttt cctcatgaa aaggcaattt  
 3961 ccacactcac aatatgcgac aaagacaaac agagaacaat taatgtgctc ctctctaattg  
 4021 tcaaaattgt agtggcaag agggacaac aatctcaagt tctgagttag ttttagtgat  
 4081 tggataaag gctttgacct gtgagctcac ctggacttca tatccttttg gataaaaagt  
 4141 gcttttataa ctttcaggtc tccagctctt tattcaatg actgttgggt tagggacaga  
 4201 ccacacaaatg aatgcctggc ataggaaagg gcagcagagc cttagctgac cttttcttgg  
 4261 gacaaagcatt gtcaacaat gtgtgacaaa actatttcta ctgctttgca cagctgtgct  
 4321 gggcagggcg atccattgac acctatccca ggtaaccttc caactgcaa agattgttg  
 4381 ctactctct ctagaaagt ctgcagact gacatgcatt tcataggtag agataacatt  
 4441 tactgggaag cacatctatc atcacaaaa gcaggcaaga ttttcagact tcttagtgg  
 4501 ctgaatataga agcaaaagac gtaattaaa acaaatgaa acaaaaaa tcaagtggata  
 4561 cctgtgtgtgt agacatccag caaaaaata ttattgac taccatcttg tcttaagtcc  
 4621 tcagacttag caaggagaat gtgatttcc acagtatata tgtttcaca aaaggaagga  
 4681 gagaacaaa agaaaatggc actgactaaa cttcagctag tggatagga aagtaattct  
 4741 gcttaacaga gattgcagt atctctatgt atgtcctgaa gaattatgtt gtactttttt  
 4801 cccccatttt taaatcaac agtgccttac agaggtcaga atggtttctt tactgtttgt  
 4861 caattctatt atttcaatc agtgccttac agaggtcaga atggtttctt tactgtttgt  
 4921 atattatgat tgtccctcga accatgaaca ctctccagc tgaatttcc aattcctctg  
 4981 tcatctgcca gggcatttaag ttattcatgg agatctttg aggaacactg caagttcata  
 5041 tcataaacac atttgaatt gagtatgtgt ttgatttga tggagctatg tttgtgtga  
 5101 tctctgaaa aaaagtgtgt tataaagcat tccacccat aaaagatag atttaatat  
 5161 tccagctata ggaagaaaag tgcgtctgct ctactctca gtctcagttg gctcctcac  
 5221 atgcatgctt ctttatttct cctattttgt caagaaaaata ataggtcacg tcttgttctc  
 5281 acttatgtcc tgcctagcat ggtcagatg caggtttag atacaagaag gatcaatga  
 5341 aacagacttc tggctgtgta cctacaacca tagtaataag cacactaact aataattgct  
 5401 aattatgttt tccatctcta aggttcccat atttttctgt tttcttaag atccattat

5

10

15

20

25

30

35

5461 ctggttgtaa ctgaagctca atggaacatg agcaatatatt ccagtccttc tctccatcc  
 5521 aacagtcctg atggattagc agaacaggca gaaacacat tgttaccag aataaaaaac  
 5581 taatatttgc tctcatttca atccaaaatg gaacctatga aactaaaatc taacccaatc  
 5641 ccataaaatg atttctatgg tgtcaaaagt caaacttctg aagggaacct gtgggtgggt  
 5701 cacaaattcag gctatatatt ccccaaggct cagccagtgg atcaatcatt catctcgtga  
 5761 ctctctcgtg tgtgtgtgtt acctatatat ctaaaatttaa tatttcgttt attaaaattt  
 5821 aatatatttc gacgatgaat ttctcaagga tatttttctt cgtgttcgct ttggttctgg  
 5881 ctttgtcaac agtttcggct gcgcagagc gaaaaggtae ccagggtgcag ctgcaggagt  
 5941 cggggggagg ctttgtaaag ccgggggggt cctttagagt ctctgtgca gcctctggat  
 6001 tcactttcag aaacgctgg atgagctggg tccggcaggc tccagggaag gggctggagt  
 6061 gggctggcgg tattaaaag aaaaatgatg gtgggacac agactatgct gcacctgtga  
 6121 aaggcagatt caccatctca agagatgatt caaaaaaac gtatatctg caaatgaata  
 6181 gctgaaagc cgaggacaca gccgtatat actgtaccac ggggattatg ataactttg  
 6241 ggggagttat ccttccccg aattggggcc aggga:ccct ggtcacctgc tctcagcct  
 6301 ccaccaaggg ccatcggtc ttccccctgg caccctctc caagagcac tctgggggca  
 6361 cagcggcctt gggctgcctg gtcaaggact acttccccga accgggtgag gtgtcgtgga  
 6421 actcaggcgc cctgaccagc ggcgtgcaca cctttccggc tgcctacag tctcaggac  
 6481 tctacttctt tagcaacgtg gtgacctgc cctccagcag ctctgggacc cagacctaca  
 6541 tctgcaactg gaatcacaa cccagcaaca ccaagggtgga caagaaagt gagccaaat  
 6601 ctgtgtgaca aactcacaca tgcceacctt gccagcacc tgaactcctg gggggaccgt  
 6661 cagcttctct cttccccca aaacccaagg acacctcat gatctccgg accctgagg  
 6721 tcaatgcgtt ggtgtggac gtgagccag aagacctga ggtcaagttc aactgggtacg  
 6781 tggacggcgt ggaagtgcac aatgccaaga caaagccgcg ggaggagcag tacaacagca  
 6841 cgtaccgtgt ggtcagctc ctacccgtcc tgcaccagga ctgggtgaat ggaayaggt  
 6901 caaagtga ggtctccaa aaagccctcc cagcccccat cgagaaaaac atctccaaag  
 6961 ccaaaaggca gcccgagaa ccacaggtgt acacctgccc ccatcccg gatgagctga  
 7021 caaagaacca ggtcagcctg acctgcctgg tcaaggctt ctatcccg gacatcgccg  
 7081 tggagtggga gagcaatgg cagccggaga acaactaaa gaccacgct ccggtgctgg  
 7141 actccgacgg ctctctctc ctctacagca agctaacgt ggacaagagc aggtggcagc  
 7201 aggggaacgt ctctctatgc tccgtgatgc atgaggtctt gcacaaccac tacacgcaga  
 7261 agagcctctc cctgtctccg ggtaaaagcgc cagagccgaa gcttctctat gactgacac  
 7321 agccaccctc ggtgtcagtg tccceaggac aaacggccag gatcacctgc tctggagatg  
 7381 cattgccaga aaatatgtt tattgtacc agcagaagtc aggccaggcc cctgtgtgtg  
 7441 tcatctatga ggacagcaaa cgaacctccg ggatccctga gagattctct ggtccagct  
 7501 cagggaacaat ggccacctt actatcagtg gggcccagggt ggaagatgaa ggtgactact  
 7561 actgttactc aactgacagc agtgggtatc atagggaggt gttcagcgga gggaccaagc  
 7621 tgaacgtcct aggtcagccc aaggctgccc cctcgtctac tctgttccca cctctctctg  
 7681 agagccttca agccaacaag gccacactgg tgtgtctcat aagtactcc taccgggag  
 7741 cctgacagt ggcctgggaag gcagatagca gccccgtcaa ggcggggagtg gagaccacca

7801 caccctccaa acaagcaac aacaagtacg cggccagcag ctacctgagc ctgacgccttg  
 7861 agcagtggaa gtcccacaaa agctacagct gccaggtcac gcatgaagg agcaccgtgg  
 7921 agaagacagt gcccctgca gaatgttcac gcaggaggga ggaaggcc cttttgaag  
 7981 ggggaggaaa cttcgcgcca tgactcctct cgtgcccccc gcacggaaca cttatgtgca  
 8041 gaggccctc tgccattgct gcttcctctg ccttcctctg tcactctgaa tgtggcttct  
 8101 ttgtactgc cacagcaaga aataaatct caacatctaa atgggtttcc tgagattttt  
 8161 caagagtctg taagcacatt ccttccccag cacccttgc tgcaggccag tgcaggccac  
 8221 caacttggct actgctgccc atgagagaaa tccagttcaa tttttccaa agcaaaatgg  
 8281 attacatatg ccttagatcc tgattaacag gtgttttcta ttatctgtgc ttctgcttca  
 8341 cccacattat cccattgctt cccctcgagg gggggcccg tacccaattc gccctatagt  
 8401 gagtgcgtat acgcgcgctt ccttgccagc catccctctt tgcgtagctg ggaaccctt  
 8461 tgtaagcgt taatatcttg ttaaaattcg cgttaattt ttgttaaatc agtccatttt  
 8521 ttaaccaata ggcgaaatc ggcgaatcc cttacaatc aaagaatatg accgagatatg  
 8581 ggttgagtgt tgttccagtt tggacaaga gtccactatt aaagaacgtg gactccaacg  
 8641 tcaaggggc aaaaaccgtc tatcagggcg atggccact actccggat catatgacaa  
 8701 taagtgtatc cacttaact taatgatttt taccaaatc attaggggat tcatcagttc  
 8761 tcaagggtcaa cgagaattaa cattccgtca ggaagcttta tgatgatgat tgccttaaaa  
 8821 acttactcaa tggctggtta tgcataatgc aatacatgcg aaaaacctaa agagcttgc  
 8881 cgataaaaaa ggcgaattta ttgctattta ccgcgctttt ttattgagct tgaagataaa  
 8941 ataaatataga taggttttat ttgaagctaa atcttcttta tcgtaaaaa tgcctctctg  
 9001 ggttataag aggttcatta tatctcgcg aataacatca ttgtgtacg aaataactaa  
 9061 gcacttgtct cctgttttact cccctgagct tgagggttta acatgaaggt catcgtatgc  
 9121 aggataataa tacagtaaaa cgttaaaaca ataatacaaa tccagccatc ccaaattgggt  
 9181 agtgaatgat tataatccctg taaagcact tgcgtatgac tctttgtttg gatagacatc  
 9241 aggtcaacca ataactccctg agcgatccca cccagcgca ataaattaa aacaggga  
 9301 actccctgta atgcaggtaa agcgatccca cccagcgca ataaattaa aacaggga  
 9361 actaaccac cttcagatat aaacgctaaa aggcataatg cactactatc tgcaataaat  
 9421 ccgagcagta ctgcccgtttt ttgcgccatt tagtggctat tcttctgcc acaaggctt  
 9481 ggaatactga gtgtaaaaa caagaccog taatgaaaag ccaaccatca tgctattcat  
 9541 catcagatt tctgtaatag caccacacog tctggtattg gctatcaatg cgtgaaata  
 9601 ataataca aatggcatcg ttaataaagt gatgtatacc gatcagcttt tgttcccttt  
 9661 agtgaaggtt aatggcgcg ttggcgtaat catggtcata gctgttctt ggtgaaatt  
 9721 gttatccgt cacaattcca cacaacatc gagccggaag cataaagt aaagcctggg  
 9781 gtccctaag agtgagctaa ctacattaa ttgcgttgcg ctactgcc gcttccagt  
 9841 cgggaacct gtcgtgccag ctgcattaat aacgcgggg agagcgggtt  
 9901 tgcgtattgg gcgtcttcc gcttctcgc tcaatgactc gctgcgctg gctgtcggc  
 10021 tgcggcagc ggtatcagct cactcaagg cgttaatacg gttatccaca gaatcagggg  
 10081

5

10

15

20

25

30

35

10141 ataacgcagg aaagaacatg tgagcaaaa ggcagcaaaa ggcaggaac cgtaaaaag  
 10201 ccgggttgct ggcgtttttc catagctcc gccccctga cagcatcac aaaaatcgac  
 10261 gctcaagtca gagtggcga aaccgacag gactataaag ataccaggcg ttccccctg  
 10321 gaagctccct cgtgcgtct cctgtccga cctgcgct tacoggatac ctgtccgct  
 10381 ttctcccttc ggaagcgtg gcgtttctc atagctcag ctgaggtat ctacgttcg  
 10441 tggaggtcgt tcgtccaaag ctgggtcgtg tgcacgaacc cccgttcag ccgaccgct  
 10501 gcgccttacc cggtaactat cgtcttgagt ccaaccgggt aagacacgac ttatcgccac  
 10561 tggcagcagc cactggtaac aggtatagca gagcgaggt ttagggcgt gctacagagt  
 10621 tcttgaagt gtcgctaac tacggctaca ctagaaggac agtattgggt atctgcgctc  
 10681 tgcgaaagc agttacctc ggaagagag ttggtagctc ttgatccgc aaacaaacca  
 10741 ccgctggtag cgggtgttt tttgttgca agcagcagat tacgcgcaga aaaaaaggat  
 10801 ctcaagaaga tcccttgatc tttctacgg ggtctgacg tcaagtgaac gaaactcac  
 10861 gtaagggat ttgggtcatg agattatcaa aaaggatctt cacctagatc ctttaaat  
 10921 aaaaatgaag ttttaaatca atctaaagta tatatagta aacttggtct gacagttacc  
 10981 aatgcttaac cagtggagga cctatctcag cgtctgtct attcgttca tccatagtgtg  
 11041 cctgactccc cgtcgtgtag ataactcga tacgggaggg ctaccatct ggcaccagt  
 11101 ctgcaatgat accgcgagc caacgctcac cggctccaga ttatcagca ataaaccagc  
 11161 cagccggaag ggcgagcgc agaagtgtc ctgcaacttt atcgcctcc atccagtcta  
 11221 ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagttg cgcacgttg  
 11281 ttgccattgc tacaggcatc gtggtgtcac gtcgtcgtt tggataggct tcaatcagct  
 11341 cgggttccca acgatcaagg cgaattacat gatccccat gtgtgcaaa aaagcgggta  
 11401 gctccttcgg tctccgac gtgtgcagaa gtaagtggc cgcagtgtta tcaatcattg  
 11461 ttatggcagc actgcataat tctcttactg tcatgcatc cgtaaagatgc tttctgtga  
 11521 ctgggtgagta ctcaaccaag tcattctgag aatagtgat gcggcgacog agttgtctt  
 11581 gccggcgctc aatacgggat aataccgac cacaatagc aactttaaaa gtgtcatca  
 11641 ttggaaaaacg ttcttcgggg cgaactct caaggatctt accgctgtg agatccagt  
 11701 cgaatgaacc cactcgtgca ccaactgat ctccagcatc tttactttc accagcgtt  
 11761 ctgggtgagc aaaaacagga aggcaaatg ccgcaaaaa ggaataaagg gcacacgga  
 11821 aatgttgaat actcactatc ttctttttc aatattattg aagcatttat cagggttatt  
 11881 gtctcatgag cggatacata tttgaatgta tttagaaaaa taacaaaaa ggggttccg  
 11941 gcacatttcc ccgaaagtg ccac

//

SEQ ID NO:99 (pTnMCS(CHOVep-prepro-HCPro-Lys-CPA))

35 1 ctgacgcgc ctgtagcgc gcattaagcg cggcggtgtg ggtggttaog cgcagcgtga  
 61 ccgtacact tgcagcgcc ctagcgccg ctctttgcg ttcttccct tctttctcg  
 121 ccaogttcgc cggatcaga ttggctattg gccattgat acgttgatc catatcataa  
 181 tatgtacatt tatattggct catgtccaac attaccgcca tgtgacatt gattattgac  
 241 tagtatttaa tagtaatcaa ttacgggggtc attagttcat agccatatata tggagttcog

301 cgttacataa cttacgggttaa atggcccgcc tggctgaccg cccaagacc cccgccatt  
 361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgaagtca  
 421 atgggtggag tatttaccgtt aaactgcca cttggcagta catcaagtgt atcatatgcc  
 481 aagtagccc cctattgacg tcaatgaagg taaatggccc gctggcatt atgcccagta  
 541 catgacctta tgggactttc ctacttgcca gtacattac gtattagta tcgtattac  
 601 catggfagat cgggtttggc agtacctcaa tggcggtgga tagcggtttg actcacgggg  
 661 attccaagt ctccaccoca ttgacgtcaa tgggagtttg tttggcacc aaatcaacg  
 721 ggaatttcca aaatgtcgtl acaactccgc ccaattgacg caaatgggag gtagcgtgt  
 781 acggtgggag gtctatataa gtagagctcg tttagtgaac cgtcagatcg cctggagacg  
 841 ccaccacgc tgttttgacc tccatagaag acaccgggac cgtccagcc tccgcgccg  
 901 ggaacgggtgc attggaacgc ggaattcccg tgcgaagagt gacgtaagta ccgctatag  
 961 actctatagg cacaccctt tggctcttat ggtctcttat cgttttttg cttggggcct  
 1021 atacacccc gcttccctat gctatagggt atggtatagc ttgacctata ggtgtgggtt  
 1081 attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac  
 1141 atgggtcttt gccacaacta tctctattgg ctatacgcca atactctgc ctccagagac  
 1201 tgacacggac tctgtatttt tacaggatgg ggtccatttt attattaca aattcacata  
 1261 tacaacaacg ccgtccccc tgcccgcagt ttttattaaa catagcgtgg gatctccacg  
 1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctcgg ttagcggcgg agcttccaca  
 1381 tccagccct ggtcccatgc ctccagcggc tcatggtcgc tggcagetc ctgtctccta  
 1441 acagtggagg ccagacttag gcacagaca atgcccacca ccaccagtgt gccgacaaag  
 1501 gccgtggcgg tagggtatgt gtctgaaat gagcgtggag attgggctcg cagggtgac  
 1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgtgtattc  
 1621 tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cgttgagggg cagtgtagtc  
 1681 tgagcagtag tegtgtctgc cgcgcgcgc accagacata atagctgaca gactaaraa  
 1741 ctgttccctt ccatgggtct tttctgcagt caccgtcggg ccatgtgga actcgatatt  
 1801 ttacacgact ctctttacca attctgccc gaattacact taaaacgact caacagctta  
 1861 acgttggctt gccacgcat acttgactgt aaaactctca ctcttacoga acttggccgt  
 1921 aactgcca ccaagcag acaaaacat aacataaac gaatcgaccg attgttaggt  
 1981 aatcgtcacc tccacaaga cgcactcgt gtatacgtt ggcattgtag ctttatctgt  
 2041 tggggcaata cgtatcccc tgtacttgtt gactgtctg atattctga gcaaaaaaga  
 2101 ctatgttat tggagcttc agtcgacta caccgtcgtt ctgttactct tcatgagaaa  
 2161 ggttcccg ttcagagca atgttcaaa gaaactcgt accaatctct agccgacctt  
 2221 ggagcattc tacogagtaa caccacacg ctcatgttca gtgatgttg ctttaaagt  
 2281 catgggtata aatccgttga gaagctgggt tggtagctgt taagtgcagt aagaggaaaa  
 2341 gtacaatatg cagacctagg agcggaaaa tggaaacta tcagcaactt acatgatag  
 2401 tactatagc actcaagac tttaggctat aagagctga ctaaaagcaa tccaactca  
 2461 tgcctaaatc tattgtataa atctgctct aaaggccgaa aaatcagcg ctgcacacgg  
 2521 actcattgtc accaccgtc acctaaatc tactcagcgt cggcaaaagg gccatgggtt  
 2581 ctagcaacta acttacctgt tgaattoga acaccaaac aacttgttaa tatctatcgg

5

10

15

20

25

30

35

2641 aagcgaatgc agattgaaga aaccttcoga gacttgaaaa gtccctgccta cggactaggc  
 2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatgcacctg  
 2761 atgcttcaac taacatgttg gcttgcggc gtccatgctc agaacaagg ttgggacaaag  
 2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg  
 2881 gaagttttgc ggcattcttg ctacacaata caaagggaag acttactcgt ggctgcaacc  
 2941 ctactagctc aaaatttatt cacacatggt tacgctttgg ggaattatg aggggatcgc  
 3001 tctagagcga tccgggatct cgggaaaagc gttggtgacc aaaggtgcct ttatcatca  
 3061 ctttaaaaat aaaaaaat tactcagtcg cgtttataag cagcaattaa ttatgattga  
 3121 tgccatcatc acaacaaaaa ctgatttaac aaatggttgg tctgccttag aaagtatat  
 3181 tgaacattat cttgattata ttattgataa taataaaaaa cttatcccta tccaagaagt  
 3241 gatgcctatc attggttga atgaacttga aaaaaattag ccttgaatac attactggta  
 3301 aggtaaacgc cattgtcagc aaattgatcc agagaaccca acttaaaagt ttctgacgg  
 3361 aatgttaatt ctggttgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaaac  
 3421 attaaagttaa ggtggataca catcttgca tatgatccg gtaattgtgag ttagctcact  
 3481 cattaggcac ccaggcttt acactttatg ctccggctc gtatgttgg tggaaattgtg  
 3541 agcggataac aatttcacac aggaacacgc tatgacctg attacgcca gcgcgaatt  
 3601 aacctctact aaagggaaca aaagtggag ctccacccgg ttggcgccg cctagaact  
 3661 agtggatccc ccgggttga ggaattgat atcaagctta tgcataccgc tgacctcag  
 3721 ctgcagaaa atgccagtg gactatgac tcaatccaa aggagcttga cctgatacct  
 3781 gatcttctc aaacagggga acaacacaa tccacaaaa cagctcagag agaaaccatc  
 3841 actgatggct acagaccaa ggtatgcaat ggcaatccat tgcacattca tctgtgacct  
 3901 gagcaaatg atttctctt ccatgaatgg ttgcttctt cctcatgaa agggcaattt  
 3961 ccacactcac aatatggac aaagacaaac agagaacaat taatgtgctc ctcccaatg  
 4021 tcaaatgtg agtggaag aggagaaca aatctcaagt tctgagttag ttttagtga  
 4081 tggataagag gctttgacct gtgagctcac ctggacttca tatccttttg gataaaaagt  
 4141 gcttttataa ctttcaggtc tccgagctct tattcatgag actgtttggt tagggacaga  
 4201 cccacaatga aatgcctggc ataggaaagg gcagcagagc cttagctgac ctttcttgg  
 4261 gacaagcatt gtcaacaat gtgtgacaaa actatttga ctgcttttga cagctgtgct  
 4321 gggcaggggc atccattgac accatccca ggtaaccttc caactgcaag agattgtttg  
 4381 ctactctct ctgaaaagct tctgcagact gacatgcatt tcataggtag agataacatt  
 4441 tacttgggaag cacatctatc atcacaaaaa gagggcaaga ttttcagact tcttagtgg  
 4501 ctgaaataga agcaaaagac gtaattaaaa acaaaaaaaa tcaatttgata  
 4561 cctgtgtgtg agacatccag caaaaaata ttatttgcac taccatcttg tcttaagtcc  
 4621 tgcagcttag caaggagaat ttagatttcc acagtatata tgttttca aagggaagga  
 4681 gagaacaaa agaaatggc actgactaaa cttagcttag tggatataga aagttaattct  
 4741 gcttaacaga gattgcagtg atctctatgt atgtcctgaa gaattatgtt gtactttttt  
 4801 ccccatctt taatcaaac agtgcctttac agaggtcaga atggtttctt tactgtttgt  
 4861 caattctatt atttcaatc agaacaatat ctctataac tgaatatat ttgctattgt  
 4921 atattatgat tgtccctcga accatgaaca ctctccagc tgaatttcac aattcctctg

5

10

15

20

25

30

35



4981 tcatctgcca ggcatttaag ttattcatgg aagatctttg aggaacactg caagttcata  
 5041 tcaataacac atttgaatt tggattggt ttgattgta tggagctatg ttttgcgtga  
 5101 tctctagaaa aaaagtctgt tataaagcat tccacccat aaaaagatag atttaatat  
 5161 tccagctata ggaagaaaag tgcgtctgct ctcaactcta gtctcagttg gctccttcac  
 5221 atgcagtctt ctttatttct cctattttgt caagaaaaata ataggtcacg tcttgttctc  
 5281 acttatgtcc tgcctagcat ggctcagatg cagcttgtag atacaagaag gatcaaatga  
 5341 aacagacttc tggctgttta cctacaacca tagtaataag cacactaact aataattgct  
 5401 aattatgttt tccatctcta aggttcccat atttttctgt tttcttaaa atccattat  
 5461 ctgggtgtta ctgaagctca atggaacatg agcaatattt cccagtcttc tctcccatcc  
 5521 aacagctctg atggattagc agaacaggca gaaacacat tgttaccacg aattaaaaac  
 5581 taatatttgc tctccattca atccaaaatg gaaccttga aactaaaac taaccaatc  
 5641 ccaataaatg atttctatgg tgtcaaaagt caaacttctg aagggaaacct gtgggtgggt  
 5701 cacaattcag gctatatatt cccagggtt cagccagttg atcaatcatt catctcgtga  
 5761 ctcttcgtg tgtggtgttt acctatatat ctaaaattaa tatttcgttt attaaaaattt  
 5821 aatatattc gacgatgaat ttctcaagga tattttctt cgtgttcgct ttggttctgg  
 5881 ctctgtcaac agtttcggct ggcacagacg cgaaggtac ccaggtgcag ctgcaggagt  
 5941 cggggggagg ctgtgtaaa cggggggggt cctcttaggt cctctgtga gcctctggat  
 6001 tcactttcag aaacgcttg atgactggtg tccgccaggc tccagggaaag gggctggagt  
 6061 gggctggcgt tattaaaagc aaaaattgat gtgggataac agactatgct gcaccctga  
 6121 aagcagact caccatctca agagtatatt caaaaaaac gtatatctg caaatgaata  
 6181 gctgaaagc cgaggacaca gccgtatat actgtaccac ggggattatg ataacatttg  
 6241 ggggagttat cctccccc cgtggggcc agggaaacct ggtcacctgc tctcagcct  
 6301 ccaaccaagg ccatacgttc ttccccttg caccctctc caagagcacc tctgggggga  
 6361 cagcggcctt gggctgctg gtcaaggact acttcccga accgggtgacg gtgtcgtgga  
 6421 actcaggcgc cctgaccagc ggcgtgcaca ctttccggc tgcctacag tctcaggac  
 6481 tctacttctt tagcaacgtg gtgacctgc cctccagcag ctggggcacc cagacctaca  
 6541 tctgcaact gaatacaag cccagcaaca ccaaggtgga caagaaagt gagcccaaat  
 6601 ctgttgaca aactcacaca tgcacacctg gccagacc tgaactctg gggggacctg  
 6661 cagtcttctt cttccccc aaaccaaag acacctcat gatctccgg accctgagg  
 6721 tcaatgcgtt ggtggtggac gtgagccacg aagacctga ggtcaagttc aactggtagc  
 6781 tggacggcgt ggaaggtgcat aatgccaaga caaagcgcg gaggagcag tacaacagca  
 6841 cgtaccgtgt ggtcagcgtc ctaccgtcc tgcaccaga ctggctgaat ggcaaggagt  
 6901 acaagtgcga ggtctccaac aagcctcc cagccccc atcctcaaa  
 6961 ccaaaaggca gcccagaga ccacaggtgt acacctgcc ccatcccg gatgagctga  
 7021 ccaagaacca ggtcagcctg acctgcctg tcaaggctt ctatccagc gacatgcgcg  
 7081 tggagtggga gacaaatgg cagcggaga caactacaa gaccagcct cccgtgctgg  
 7141 actccgacgg ctcttcttc ctctacaga agctaccgt ggacaagagc aggtggcagc  
 7201 agggggaagc ctctctatgc tccgtgatgc atgaggtctt gcacaaccac tacacgaga  
 7261 agagcctctc cctgtctcog ggtaagcgc cagagccgaa aaagctttcc tatgagctga

5

10

15

20

25

30

35

7321 cacagccacc ctccggtgtca gtgtcccccag gacaaacggc caggatcacc tgcctcggag  
 7381 atgcattgcc agaaaaatat gtttatgggt accagcagaa gtcaggccag gccctcgtgg  
 7441 tggatcatcta tgaggacagc aaacgacct cgggatccc tgagagattc tctggtcca  
 7501 gctcaggagc aatggccacc ttgactatca gtggggccca ggtggaagat gaaggtgact  
 7561 actactgtta ctcaactgac agcagtgggt atcatagga ggtgttcagc ggaggacca  
 7621 agctgaccgt cctaggtcag cccaaggctg cccctcgggt cactcgttc ccacctcct  
 7681 ctgaggagct tcaagccaac aaggccacac tgggtgtgtc cataagtgc tcctaccgg  
 7741 gagccgtgac agtgccctgg aaggcagata gcagccccgt caaggcggga gtggagacca  
 7801 ccacaccctc caaacaagc aacaacaagt acgaggccag cagctacctg agcctgacgc  
 7861 ttgagcagtg gaagtccac aaagctaca gctgccaggt cacgaatgaa gggagcaccg  
 7921 tggagaagac agtgccctc scagaaatgtt caccggggag ggaaggaaagg gccctttttg  
 7981 aagggggagg aaacttcgag ccatgactcc tctcgtgccc ccgacaggga acactgatgt  
 8041 gcagagggcc ctctgccatt gctgcttctt ctgcccttcc tgcactctc gaatgtggct  
 8101 tctttgtctac tgccacagca agaaataaaa tctcaaatc taaatgggtt tctgagatt  
 8161 ttcaagagct cgttaagcac attccttccc cagcaaccc tgcgcaggc cagtgcagg  
 8221 caccaacttg gctactgctg cccatgagag aaatccagtt caatatattt caaagcaaaa  
 8281 tggattacat atgccctaga tcttgattaa caggtgtttt gtattatctg tgccttcgct  
 8341 tcaccacat tatccattg cctccctcag agggggggcc cggtaaccaa ttgccttat  
 8401 agtgaagtcgt attagcgcg ctcactggcc gtggttttac aagtcgtga ctgggaaaaac  
 8461 cctggcgcta cccaaactaa tgccttgca gcacatccc ctttcggcag ctggcgtaat  
 8521 agcgaagagg ccgcaccga tgccttccc caacagttgc gcagcctgaa tggcgaatgg  
 8581 aaattgtaag cgttaatat ttgttaaaat tgcgtttaa tttttgttaa atcagctcat  
 8641 tttttaacca ataggccgaa atcggcaaaa tcccttataa atcaaaagaa tagaccgaga  
 8701 tagggttgag tgttgtcca gtttggaaca agagtccact attaaagaa gtggactcca  
 8761 acgtcaagg gcgaaaaacc gtctatcagg gggatggccc actactccgg gatcatatga  
 8821 caagatgtgt atccacctta acttaatgat ttctaccaaa atcattaggg gattcatcag  
 8881 tgcacagggt caacgagaat taacattccg tcagggaagc ttatgatgat gatgtgctta  
 8941 aaaacttact caatggctgg ttatgcataat cgcaatacat gcgaaaaacc taaaagagct  
 9001 tgcgataaa aaaggccaat ttattgctat ttaccggggc tttttattga gcttgaaaga  
 9061 taaataaaat agatagggtt tatttgagc taaatcttct ttatcgtaaa aaatgccctc  
 9121 ttgggttattc agaggggtca ttatatctc gggaataaca tcaattgggtg acgaataaac  
 9181 taaggacttg tctcctgttt actccctga gcttgagggg ttaacatgaa ggtcatcgat  
 9241 agcaggataa taatacagta aaacgctaaa ccaataatcc aaatccagcc atcccaatt  
 9301 ggtagtgaat gattataaat aacagcaaac agtaatgggc caataacacc ggttgcattg  
 9361 gtaaggctca ccaataatcc ctgtaaagca ccttgctgat gactctttgt ttggatagac  
 9421 atcaatccct gtaatgcagg taaagcgatc ccaccaccag ccaataaaat taaaacaggg  
 9481 aaaactaacc aaccttcaga tataaacgct aaaaaggcaa atgcactact atctgcaata  
 9541 aatccagaca gtactgccgt ttttcgccc atttagtggc tatcttctcc gccacaaagg  
 9601 ctggaatac tgagtgttaa agaccaagac ccgtaatatgaa aagccaacca tcatgctatt

5

10

15

20

25

30

35

9661 catcatcag atttctgttaa tagcaccaca cgtgctgga ttggtatca atgcgctgaa  
 9721 ataataatca acaaatggca tegttaaata agtgaatgat accgatcagc ttttgttccc  
 9781 tttatgtagg gttaatgagc cgtttggcgt aatcatggct atagctgttt cctgtgtgaa  
 9841 attgttatcc gtcacaatt ccaacaaca tagagccgg aagcataaag tgtaaagcct  
 9901 ggggtgccta atgagttagc taactacat taattgctt cgtctcaatg ccgcgtttcc  
 9961 agtcgggaaa cctgtcgtgc cagctgcatt aatgaatcgg ccaacgcgcg gggagaggcg  
 10021 gtttgggtat tggggtctct tccgtctct cgtcactga ctcgtgcgc tccgtgttcc  
 10081 ggtctgggag agcgttatca gctcactcaa agcggtaat acggttatcc acagaatcag  
 10141 gggataacgc aggaagaac atgtgagcaa aagggccagg aaagccagg aaccgtaaaa  
 10201 aggcgcgtt gctggcgttt ttccataggc tccgcccccc tgacgagcat cacaaaaatc  
 10261 gacgtcaag tcagaggtgg cgaaccocga caggactata aagataccag gcgtttcccc  
 10321 ctggaagctc cctcgtgcgc tctcctgttc cgacctcgc gcttaccga tacctgtccg  
 10381 cgtttctccc ttcgggaagc gtggcgcttt ccatagctc agctgtagg tatctcagtt  
 10441 cgggttaggt cgttcgtccc aagctgggt gtgtgcagc acccccggt cagcccgacc  
 10501 gctggcctt atccggtaac tatcgtcttg agtccaaacc ggtaaagacac gacttatcgc  
 10561 cactggcagc agccactggt aacaggatta gcagagcag gtatgtagg ggtgctacag  
 10621 agttcttgaa gtggtggcct aactacggct aactagaaag gacgtatctt ggtatctgog  
 10681 cctcgtgaa gccagtacc ttcggaaaaa ggttggtag ccttctgacc ggcaaacaaa  
 10741 ccacgcgtg tagcgtggt tttttgttt gcaagcagca gattacgcgc agaaaaaaag  
 10801 gatctcaaga agatcctttg atcttctta cggggtctga cgtcagtggt aacgaaaaact  
 10861 cagcttaagg gatttgggtc atgagattat caaaaaagat ctccacctag atccttttaa  
 10921 attaaaaatg agtttttaa tcaatctaaa gtatatatga gtaaacctgg tctgacagtt  
 10981 accaatgctt aatcagttag gcacctatct cagcgatctg tctatttctg tcatccatag  
 11041 ttgctgact cccgctcgtg tagataacta cgatacggga gggcttacc tctgggcrcca  
 11101 gtgctgcaat gatacgcga gaccacgct caccggctcc agatttatca gcaataaacc  
 11161 agccagcccg aaggcccgag cgcagaagtg gtcctgaac ttatccgcc tccatccagt  
 11221 ctattaattg ttgcccggaa gctagagtaa gtatttgcg gtttggtagt gcttcattca  
 11281 ttgttgccat tgctacaggc atcgtggtgt cagcgtcgtc gtttggtagt gcttcattca  
 11341 gctccgggtc ccaacgatca aggcaggtta catgatccc catgtgtgac aaaaaagcgg  
 11401 ttagctcctt cggctcctcg atcgttgtca gaagtaagtt ggcgcagtg ttatcactca  
 11461 ttgttatggc agcactgcat aattctctta ctgtcatgac atccgtaaga tgcctttctg  
 11521 tgactgggtga gtactcaac aagtcattct gagaatagt tatgcccga cggagtgtct  
 11581 ctgcccgcgc gtcaatacgg gataatccg cgccacatag cagaacttta aaagtgtcga  
 11641 tcaattgaaa acgttcttcc ggcgaaaaa tctcaaggat ctaccgctg ttgagatcca  
 11701 gttcgtatga acccactcgt gcacccaact gatcttcagc atcttttact ttaccagcg  
 11761 ttctctgggtg agcaaaaaa ggaaggcaaa atgcccgaat aaagggaata agggcgacac  
 11821 ggaatgttg aatactcata ctctctctt ttcaatatta ttgaagcatt tatcagggtt  
 11881 attgtctcat gagggtatc atatttgaat gtatttagaa aaataaaaa atagggggttc  
 11941 cggcacatt tccccgaaa gtgccac

SEQ ID NO:100 (pTnMCS (CMV-prepro-HCPPro-CPA))

1 ctgacgccc ctgtagcggc gaattaagcg cggcgggtgt ggtggttac cgcagcgtga  
61 ccgctacact tgcacgccc ctgagcggc ctcctttcgc ttcttccct tccittctcg  
121 ccacgttcgc cggcatcaga ttggtctattg gccattgcat acgttgatc catatcataa  
181 tatgtacatt tatattggt catgtccaac attacgcca tgttgacatt gattattgac  
241 tagttattaa tagtaataca ttacgggggtc attagttcat agccatata tggagttccg  
301 cgttacataa ctacggtaa atggcccgc tggctgacg cccaaagacc ccgcccatt  
361 gacgtcaata atgacgtatg ttcccatagt aacgcaata gggactttcc attgacgtca  
421 atgggtggag tattttacggt aaactgccc catcaagtgt atcatatgcc  
481 aagtaacccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta  
541 catgacetta tgggactttc ctacttgga gtacattcac gtattagta tgcctattac  
601 catggtgatg cggttttggc agtatacaca tggcgttga tagcgtttt actcacgggg  
661 atttccaagt ctccaccca ttgacgtcaa tgggaactttg ttttggacc aaaatcaacg  
721 ggaactttcca aaatgtcgt acaactccg cccaltgacg caaatggcg gtaggcgtgt  
781 acggtgggag gtctatataa tccatagaag acacgggac cgatccagcc tccgcggcg  
841 ccatccacgc tgttttgacc tccatagaag acacgggac cgatccagcc tccgcggcg  
901 ggaacgggtgc attggaacgc ggaattcccg tgccaaagt gacgtaagta cgcctatag  
961 actctatagg cacacccct tggctcttat gcattgctata ctgtttttgg cttggggcct  
1021 atacacccc gcttctctat gctatagggt atggtatagc ttacgtata ggtgtgggtt  
1081 attgacatt attgaccact cccctatttg tgacgtatct tccattact aatccataac  
1141 atggctcttt gccacaacta tctctatttg ctatatgcca atactctgt cttcagagac  
1201 tgacacggac tctgtatttt tacaggatgg ggtccattt attattaca aattcacacg  
1261 tacaacacg cgtcccccg tgcgcgcagt ttttattaa catagcgtgg gatctccaag  
1321 cgaatctcgg tgcgtgttc cggacatggg ctcttctcgg gtagcgggg agttccaca  
1381 tccgagccct ggtcccatgc ctcacgccc gcacagcaca atgccacca ccaccagtgt gcgcacaag  
1441 acagtggagg ccagacttag gcacagcaca atgccacca ccaccagtgt gcgcacaag  
1501 gccgtggcgg tagggatagt gtctgaaat gacgtggag attgggtcgc cagggctgac  
1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgtgtattc  
1621 tgataagagt cagaggtaac tccggttgcg gtctgttaa cgggtggagg cagtgtagtc  
1681 tgagcagtag cgttgcgtgc cgcgcgcgc accagacata atagtgaca gactaacaga  
1741 ctgttctctt catgggtct tttctgcagt caccgtcggg ccatgtcga actcgatatt  
1801 ttacacgact ccttttacca attctgccc gaattacact taaaagact caacagctta  
1861 acgttgggtt gccacgcatt acttgactgt aaactctca ctttaccga acttggccgt  
1921 aacctggcaa ccaagcggag acaaaacat aacatcaaac gaatcgacc attgttaggt  
1981 aatcgtcacc tccacaaaga cgcactcgt gtatacgtt ggcattgtag ctttatctgt  
2041 tggggcaata cgtggccat tgtacttgtt gactggtctg atattcgtga gcaaaaacga  
2101 cttatgggat tggagcttc agtcgacta caggtcgtt ctgttactct ttatgagaa  
2161 ggttcccg tttcagagca atgttcaag aagctcatg accaatttct agccgacctt

2221 gggagcattc taccgagtaa caccacacgg ctcatgtgca gtgatgtgg ctttaaagtg  
 2281 ccatggata aatccgttga gaagctgggt tggtaactgg taagtcgagt aagaggaaaa  
 2341 gcataatg cagacctagg agcgaaaaac tggaaacctt t'agcaactt acatgatatg  
 2401 tcatctagtc actcaaatg tttaggctat aagagggctga ctaaaagcaa tccaatctca  
 2461 tgccaaatc tattgtataa atctcgctct aaggcgcaa aaatcagcg ctcgacacgg  
 2521 actcatgtc accaccgtc acctaaaac tactcagcgt cggcaaggga gccatgggtt  
 2581 ctacgcaacta acttacctgt tgaatttoga acocccaaac aactgtttaa tatctattcg  
 2641 aagcgaatgc agattgaaga aactctccga gacttgaaaa gtcctgccta cggactaggc  
 2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgtctgt aatcgccctg  
 2761 atgcttaac taacatgttg gcttcgggc gttcatgctc agaacaagg ttgggacaag  
 2821 cacttcacgg ctaacacagt cagaatcga acgtactct caacagttcg cttaggcattg  
 2881 gaagtcttgc ggcattcttg ctacacaata caaggggaag acttactcgt ggctgcaacc  
 2941 ctactagtc aaaaattatt cacacatggt tacgctttgg ggaatttatg aggggatcgc  
 3001 tctagagcga tccgggatct cgggaaaagc gttggt'jacc aaggtgcct ttatcatca  
 3061 ctttaaaat aaaaaacaat tactcagtcg ctgtlataag cagcaattaa ttatgattga  
 3121 tgcctacatc acaacaaaaa ctgattttaa aatgggttgg tctgccttag aaagtatat  
 3181 tgaacattat cttgattata ttattgataa taataaaaaa cttatoccta tccaagaagt  
 3241 gatgcctatc attggttggga atgaacttga aaaaaattag ccttgaatac attactggta  
 3301 aggtcaacgc cattgtcagc aaattgatcc agagaaacca acttaaaagt ttcctgacgg  
 3361 aatgttaatt ctcgcttgacc ctagacactg atgaatcccc taatgatctt ggtaaaaatc  
 3421 attaatgtaa ggtggataca catcttgta t'atgatcccg gtaatgtgag ttagctcact  
 3481 cattaggcac cccaggcttt acactttatg ctccggctc gtatgttgg tggatgttg  
 3541 agcggataac aatttcacac aggaacacagc tatgaccatg attacgcaa gcgcgaatt  
 3601 aacctcact aaagggaaca aaagctggag ctccacggg gtggcgccg ccttagaact  
 3661 agtggatccc cgggctgca ggaattcgat atcaagctta t'cgataccgc tgacctcgag  
 3721 catcagattg gctattggcc attgcatacg ttgtatccat atcataatat gtacatttat  
 3781 attggctcat gtccaaactt accgccatgt t'gacattgat tatgactag ttat'aaatag  
 3841 taatcaatta cggggtcatt agttcattgc ccataatatg agtcccgct tacataaact  
 3901 acggt'aaatg gcccgcttgg ctgaccgcc aacgaccccc gccattgac gtcaataatg  
 3961 acgtatgttc ccatagtaac gccaataggg actttccatt gacgtcaatg ggtggagtat  
 4021 ttacggtaaa ctgcccactt ggcagtaaat caagtgtatc atatgtcaag tacgcccc  
 4081 attgacgtca atgacggtaa atggcccgcc tggcattatg ccaagtacat gaccttatgg  
 4141 gactttccta cttggcagta catctacgta ttatgcatcg ctattaccat ggtgatgcgg  
 4201 ttttggcagt acatcaatgg gcgtggatag cgggttgact cagggggatt tccaagtctt  
 4261 cccccattg acgtcaatgg gagt'gtt'gtt tggcaccaaa atcaacggga ctttccaaaa  
 4321 t'gtcgt'aaac actccgcccc attgacgcaa atggggcgta ggcgtgtacg ttgggaggtc  
 4381 tatataagca gagctogttt agtgaacgt cagatcgctt ggagacgcca tccacgctgt  
 4441 tttgacctoc atagaagaca cgggacoga tccagctcc gcggccggga acggtgcatt  
 4501 ggaacgcgga tccccgtgc caagagtgc cctatagact ctataggcac

5

10

15

20

25

30

35

4561 acccttttgg ctcttatgca tgcatactg ttttgggtt ggggcctata caccocgct  
 4621 tccttatgct ataggtgatg gtatagctta gcctataggt gtgggttatt gaccattatt  
 4681 gaccactccc ctattgggtga cgatactttc cattactaat ccataacatg gctctttgac  
 4741 acaactatct ctattggcta tatgccaata ctctgtcctt cagagactga caggactctt  
 4801 gttatttttac aggatgggtt ccatttatt atttacaat tcacatatac aacaagccg  
 4861 tcccccgctc ccgcagtttt tattaacat agcgtgggat ctccacgga atctcggta  
 4921 cgtgttcggg acatgggttc ttctcggga gcgggggagc ttccacatcc gagccctggt  
 4981 cccattagca cagcgaatg tggctgctcg gcagctcctt gctcctaaca gtggaggcca  
 5041 gacttaggca cagcgaatg cccaccacca ccagctgtgc gcacaaggcc gtggcggtag  
 5101 ggtatgtgtc tgaaaatgag cgtggagatt gggctcgcac ggctgacgca gatgggaagac  
 5161 ttaaggcagc ggagagaaga gatgcaggca gctgagttgt tgtattctga taagagtcat  
 5221 aggttaactcc cgttcgggtg ctgttaacgg tggaggggcag tgtagtctga gcagtactcg  
 5281 ttgctgcgcg gcgcgccacc agacataata gctgacagac taacagactg ttcctttcca  
 5341 tgggtctttt ctgcagtcac cgtcggatca atcatratc tcgtgacttc ttcgtgtgtg  
 5401 gtgtttacct atatatctaa tttaatat ttcttactta aaatttaata tatttcgacg  
 5461 atgaatttct caaggatat ttcttctgtg ttctgttggg ttctggcttt gtcaacagtt  
 5521 tcggctgcgc cagagccgaa aggtaccacg gtgcagctgc aggagtcggg gggaggcttg  
 5581 gtaagccggg ggggtctcct tagagtctcc tgtcagcct ctggattcac ttccagaaac  
 5641 gcttgatga gctgggtccg ccaggtctca ggaagggggc tggagtgggt cggcgtatt  
 5701 aaaaacaaa ttgatgtgtg gacaacagac tatgctgcac ccgtgaaagg cagattcacc  
 5761 atctcaagag atgattcaaa aaacacgtta tatctgcaa tgaatagcct gaaagccgag  
 5821 gacacagcgt tatattactg taccacgggg attatgataa catttggggg agttatccct  
 5881 ccccgcaatt ggggccaggg aacctgggtc accgtctctt cagcctccac caaggggcca  
 5941 tgggtcttcc ccttggcacc ctcttccaag agcactctgt ggggcacagc gggcctgggc  
 6001 tgcctgttca agactactt cccgaacgg gtgacgggtgt cgtggactc aggcgccctg  
 6061 accagcggcg tgcacacctt tccggctgtc ctacagtcct caggactcta ctctcttagc  
 6121 aacgtgttga cgtgcccctc cagcagcttg ggcacccaga cctacatctg caacgtgaat  
 6181 cacaagccca gaacaccaa ggtggacaag aaagttagc ccaaatcttg tgacaaaact  
 6241 cacacatgcc caccgtgcc agcactgaa ctcttggggg gaccgtcagt ctctctcttc  
 6301 ccccccaaac ccaaggacac cctcatgac tcccgaccc ctgaggtcac atgctgtgtg  
 6361 gtggacttga gccacgaaga ccttgaagtc agttcaact ggtacgtgga cggcgtggag  
 6421 tgcataatg ccaagacaaa gcgcggggag gacagtlaca acagcagta cctgtgtgct  
 6481 agcgtctca cgtctctgca caggacttg ctgaatggca aggagtacaa gtgcaaggct  
 6541 tccaacaaag cctctccag ccccatcgag aaaacatct ccaagccaa agggcagccc  
 6601 cgagaaccac aggtgtacac cctgccccca tccgggatg agctgaccaa gaaccaggtc  
 6661 agcctgaact gcttggtaaa aggttcttat ccagcgaca tcgcgtgga gtgggagagc  
 6721 aatggcagc cggagaacaa ctacaagacc acgcctccc tgctggactc cgacggctcc  
 6781 ttcttctctt acagcaagct caccgtggac aagagcaggt ggcagcaggg gaacgtcttc  
 6841 tcatgtctcg tgatgcatga ggtctgtcac accactaca cgcagaagag cctctccctg

5

10

15

20

25

30

35

6901 tctcgggta aagcgccaga gccgaagctt tctatgagc tgacacagcc accctcgggtg  
 6961 tcagtgctcc caggcaaac gccagatc acctgctctg gagatgcatt gccagaaaaa  
 7021 tatgtttatt ggtaccagca gaagtcaggc caggccctctg tgggtgcat ctatgaggac  
 7081 agcaaacgac cctcgggat ccttgagaga ttctctggct ccagctcagg gacaatggcc  
 7141 acctgacta tcagtgggac ccaggtggaa gatgaagggt actactactg ttactcaact  
 7201 gacagcagtg gttatcatag ggaggtgttc agcggaggga ccaagctgac cgtcttaggt  
 7261 cagcccaagg ctgccccctc ggtcactctg ttccacacct cctctgagga gtttcaagcc  
 7321 aacaaggcca cactggtgtg tctcataagt gactcctacc cgggagccgt gacagtggcc  
 7381 tggaaggcag atagcagcc cgtcaaggcg ggagtggaga ccaccacacc ctccaaacaa  
 7441 agcaacaaca agtcagcgc cagcagctac cgttgagctga cgttgagca tggaagtcc  
 7501 cacaagaagt acagtgccca ggtcagcat gaaggagca ccgtggagaa gacagtggcc  
 7561 cctgcagaat gttcaccgag gaggaggga agggcccttt ttgaaggggg aggaacttc  
 7621 ggccatgac tctctctgtg ccccccgcac ggaacactga tgtgagagg gccctctgcc  
 7681 attgctgctt cctctgacct tctctgtcac tctgaatgtg gcttctttgc tactgccaca  
 7741 gaagaaata aaatctcaac atctaaatgg gtttctgtg attttcaag agtcgttaag  
 7801 cacatccct cccagcacc ccttgctgca ggccagtgcc aggcaccaac ttggctactg  
 7861 ctgcccata gagaaatcca gttcaatatt ttccaaagca aaatggatta catatgcctt  
 7921 agatccctgat taacaggtgt ttgtattat ctgtgctttc gcttaccoca cattatccca  
 7981 ttgctctccc tcgagggggg gcccggtacc caattcgccc tatagttagt cgtattacgc  
 8041 gcgctcactg gcgctgcttt tacaactgag tgaactggaa aacctggcg ttaccacact  
 8101 taatgcctt gacgacatc ccccttgc cagctggcgt aatagcgaag aggcgcgcac  
 8161 cgatcgacct tcccaacagt tggcagacct gaatggcgaa tggaaattgt aagcgttaat  
 8221 attttgttaa aattcgctt aaattttgt taaatcagct catttttaa ccaataggcc  
 8281 gaaatcgga aatccctta taaatcaaaa gaatagaccg agatagggtt gagtgtgtt  
 8341 ccagtttggg acaagagtc cactactac cgggactata tgacaagatg tgtatccacc  
 8401 accgtctatc agggcgatgg cccactactc cgggactata tgacaagatg tgtatccacc  
 8461 ttaactaat gatttttacc aaatcatta ggggattcat cagtgtcag ggtcaacgag  
 8521 aattaacatt ccgtcaggaa agcttatgat gatgatgtgc ttaaaaaactt actcaatggc  
 8581 tggttatgca tatcgcaata catcgcaaaa acctaaaaga gcttgccgat aaaaaaggcc  
 8641 aatttattgc tatttaccgc ggttttttat tgagcttgaa agataaataa aatagatagg  
 8701 ttttatttga agctaaatct tctttatcgt aaaaaatgcc ctctgggtt atcaagaggg  
 8761 ttattatatt tcgcggaata acatcatattg gtgacgaat aactaagcac ttgtctctg  
 8821 tttatccccc tgagcttgag ggggttaacat gaaggtcacc gatagcagga taataataca  
 8881 gtaaacgct aaaccaataa tccaaatcca gccatcccaa attggttagt aatgattata  
 8941 aataacagca aacagtaatg ggccaataac accggttgca ttggtaggc tcaccaataa  
 9001 tccctgtaaa gcacctgct gagcactctt tgtttggata gacatcact cctgtaatgc  
 9061 aggtaaagcg atccccacc cagccaataa aattaaaaa gggaaaaacta accaaccttc  
 9121 agataaaac gtaaaaaagg caaatgcact actatctgca ataatccga gcagtactgc  
 9181 cgttttttcg ccatctagt ggctattctt cctgccaaca aggcttggaa tactgagtgt

5

10

15

20

25

30

35

9241 aaaagaccac gaccggtgct gaaaagccaa ccatctgct atctcatc acgatttctg  
 9301 taatagcacc acacgtgct ggaattggcta tcaatgcgct gaaataataa tcaacaaatg  
 9361 gcatcggttaa taagtgtatg tataccgac agcttttctt ccttttagtg aggtttaatt  
 9421 gcgcgcttgg cgtaatcatg gtcatagctg ttccctgtgt gaaattgtta tccgctcaca  
 9481 attccacaca acatacagc cggagcaca aagtgtaaag cctgggtgct ctaatgagtg  
 9541 atgcaactca cattaatgc gttgogctca ctgcccgtt tccagtcggg aaacctgtcg  
 9601 tgcagctgc attaatgaat cggccaacgc gcgggagag gcggtttgctg tattgggctc  
 9661 tcttccgctt cctcgctcac tgactcgctg cgtcggtcg ttccggtgctg gcgagcggtg  
 9721 tcagctcact caaaggcgtt aatacggtta tccacagaat caggggataa cgcaggaaaag  
 9781 aacatgtgag caaaaggcca gaaaaggcc aggaacgta aaaaggcgc gttgctggcg  
 9841 tttttccata ggtccgccc cctgacgag catcaaaaa atcgacgctc aagtcagagg  
 9901 tggcgaaaacc cgacaggact ataaagatac caggcgcttc cccctggaag tccctcgtg  
 9961 cgctctcctg ttccgacct gccgcttacc ggatacctgt ccgctttct ccttcggga  
 10021 agcgtggcgc ttctctatag ctacgctgt aggtatctca gttcgggtgta ggtcgttgc  
 10081 tccaaagctgg gctgtgtgca cgaaccccc gttcagccc accgctgccc ctatccggt  
 10141 aactatcgtc ttgagtcaca ccggttaaga cagcacttat cgcactggc agcagccact  
 10201 ggtaaacagga ttgacagagc gaggtatgta ggcgtgtcta cagagtctct gaagtggctg  
 10261 cctaactacg gctacactag aaggacagta ttggtatct cgcctctgct gaagccagt  
 10321 accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccg tggtagcgtt  
 10381 ggtttttttg ttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct  
 10441 ttgactcttt ctacgggttc tgacgctcag tggaaacgaa actcacgtta agggattttg  
 10501 gtcattgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa atgaagtctt  
 10561 aaatcaatct aaagtatata tgagtaaaact tgggtgtgaca gttaccaatg cttaatcagt  
 10621 gaggaacctc tctcagggat ctgtctattt cgttcatcca tagttgcttg actcccgctc  
 10681 gtgtagataa ctacgatacg ggagggttta ccactggcc ccagtgtgc aatgataccg  
 10741 cgagaccacac gctcacggc tccagattta tccagataa accagccagc cggaaaggcc  
 10801 gagcgagaa gtggtcctgc aactttatcc gcctccatcc agtctattaa ttgttgcgg  
 10861 gaagctagag taagtgttc gccagttaac agtttgcga acgttgttg cattgtaca  
 10921 ggcatcgttg tgtcacgctc gtcgttttgt atggcttcat tcagctccg ttcccaaga  
 10981 tcaaggcgag ttacatgac cccatgttg tgcacaaaag cggttagctc ctccgctcct  
 11041 ccgattcgtt tcagaagtaa gttggccgca gtgttatcac tcatgggttat ggcagcactg  
 11101 cataattctc ttactgtcat gccatcgtta agatgctttt ctgtacttg tgagtactca  
 11161 accaagtcat tctgagaata gtgtatgcgg gcaccaggtt gctctgccc ggcgtcaata  
 11221 cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcttg aaacgttct  
 11281 tcggggcgaa aactctcaag gatctaacg ctgttagat ccagttcgat gtaaccact  
 11341 cgtgacacca actgatctc agcatctttt actttacca gcgtttcttg gtgacaaaa  
 11401 acaggaaggc aaatgccgc aaaaaaggga ataaaggcga caggaaaatg ttgaatactc  
 11461 atactcttcc ttttcaata ttattgaagc atttatcagg gttattgtct catgagcgga  
 11521 tacatatttg aatgtattta gaaaaataa caaatagggg ttccgcgcac atttcccca

5

10

15

20

25

30

35



11581 aaagtgccac

SEQ ID NO:101 (pTmMCS (CMV-prepro-HCPro-Lys-CPA))

1 ctgacgccc ctgtagcggc gcaatgaagc cggcgggtgt ggtggttacg cgcagcgtga  
 5 61 ccgctacact tgcagcggc ctacgcccgc ctcttttcgc ttcttccct tcctttctcg  
 121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catataataa  
 181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac  
 241 tagttattaa taglaatcaa ttacgggggc attagttcat agccatatata tggagttccg  
 301 cgttacataa cttacggtaa atggcccgc ttgctgacgc cccaacgacc ccgcccatt  
 361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca  
 421 atggtcgagg cttattcggg aaactgccc cttggcagta cateagtgat atcatatgcc  
 481 aagtagcccc cctattgacg tcaatgacgg taaatggccc gcttggcatt atgcccagta  
 541 catgacctta tgggaccttc ctacttgcca ttacatctac gtattagtcg tcgtattac  
 601 catggtgatg cgggttttggc agtacatcaa tggcgr'gga tagcggtttg actcacgggg  
 661 atttccaagt ctccaccca ttgacgtcaa tgggajtttg ttttggcacc aaatcaacg  
 721 ggaatttcca aaatgtcgtg acaactcgc ccaattgacg caaatgggoc gtagcgtgt  
 781 acggtgggag gtctatataa gcagagctcg ttagtgaac cgtcagatcg cctggagacg  
 841 ccattccacg tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcgggcg  
 901 ggaacgggtg attggaacgc ggaattcccg tgcacaagat gacgtaaagta ccgctatag  
 961 actctatagg cacaccctt tggctcttat ggtgctata ttgttttttg cttggggcct  
 1021 atacacccc gcttccttat gctataggtg atggtatagc ttgacctata ggtgtgggtt  
 1081 attgaccatt attgaccatt cccctattgg tgaogatact ttccattact aatccataac  
 1141 atggctcttt gccacaacta tctctattgg ctatatgcca atactctgc cttcagagac  
 1201 tgatacggac tctgtatttt tacaggatgg ggtcccattt attatttaca aatccacata  
 1261 tacaacaacg ccgtcccccg tgcccgcagt ttttattaaa catagcgtgg gatct'caag  
 1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctcgc gtagcggcgg agcttccaca  
 1381 tccagaccct ggtcccatgc ctccagcggc tcatggctgc tgggcagctc cttgctccta  
 1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gcgcacacag  
 1501 gccgtggcgg taggtatgt gtctgaaat gagcgtggag attgggctcg cabggctgac  
 1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtattc  
 1621 tgataagagt cagaggtaac tcccgctgag gtgctgttaa cgggtggaggc cagtgtagtc  
 1681 tgagcagtag tctgtgtgc cgcgcgcgc accagacata atagctgaca gactaacaga  
 1741 ctgttctctt ccatgggtct tttctgcagt caccgtcgga ccatgtgcga actcgatatt  
 1801 ttacacgact ctctttacca attctgccc gaattacact taaaacgact caacagctta  
 1861 acgttggtct gccacgcatt acttgactgt aaactctca ctcttaccga acttggccgt  
 1921 aacttgccaa ccaagcag agacaaacat aacataaac gaatcgaccg attgttaggt  
 1981 aatcgtccac tccacaaaga ggaactcgt gtatacgtt ggcattgtag ctttatctgt  
 2041 tgggggcaata cgatgcccc tgtacttgtt gactgtctgt aratttgtga gcaaaaacga  
 2101 cttatgggat tgcagacttc agtcgacta cagggtcgtt ctgttactct ttatgagaaa

2161 gcgtttccgc ttccagagca atgttcaaa aaagtcctag accaattct agcgacctt  
 2221 gcgagcattc taccagata caccacaccg ctcatgtca gtgatgtcg cttbaagt  
 2281 ccaggtata aatcgttga gaagctgggt tggactgggt taagtcagc aagagga  
 2341 gtacaatatg cagacctagg agcggaacac tggaaacctc tcagcaactt acatgat  
 2401 tcatctagtc actcaaacg tttaggctat aagaggctga ctaaaagcaa tccaatctc  
 2461 tgccaaattc tattgtata atctcgtct aaaggccgaa aaatcagcg ctgcacg  
 2521 actcattgtc accacocgt acctaaatc tactcagct cgcaaaagga gccatgggt  
 2581 tagcaacta acttacctgt tgaattcga acacccaaac aactgtttaa tatctattc  
 2641 aaggaatgc agattgaaga aaccttcga gacttgaana gctctgcta cggactagg  
 2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcctgtcgt aatcgcc  
 2761 atgcttcaac taacatgttg gcttgcggc gttcatgtc agaacaagg ttggaca  
 2821 cacttccagg ctaacacagt cagaaatcga aacgtactc caacagttc cttaggcat  
 2881 gaagtcttc gccattcttg ctacacata acaagggaag acttactcgt ggtgcaac  
 2941 ctactagtc aaatttatt cacacatggt tacgcttgg gaaattatg agggatc  
 3001 tctagagcga tccgggatct cgggaaagc gttggagacc aaaggtgct ttatcatc  
 3061 ctttaaaatc aaaaacaat tactcagtc ctgttataa aatggttgg tctgccttag  
 3121 tgcctacatc acaaaaaa ctgatttaac aatggttgg tctgccttag aagtatatt  
 3181 tgaacattat ctgtattata ttattgata taataaaac cttatcccta tccaaga  
 3241 gatgcctatc attggttga atgaactga aaaaaattag cctgaatac attactgg  
 3301 aggtaaacgc catgtcagc aaattgatcc aagagaacca acttaagct tctctgac  
 3361 aatgttaatt ctctgtgacc ctgagcactg atgaatcccc taatgatattt ggtaaaa  
 3421 ataaagttaa ggtggatata catctgtca tatgatccg gtaatgtgag ttagctc  
 3481 cattaggcac cccaggcttt acactttatg ctccggctc gtatgttgg tggaaattg  
 3541 agggataaac aatttcacac aggaacacgc tatgacctg attacgcaa gcgagcatt  
 3601 aacctcact aaagggaaca aaagctggag ctccacgag gtggcgccg cctagaact  
 3661 agtggatccc cgggctgca ggaattcgat atcaagctta tcgataccg tgacctgag  
 3721 catcagattg gctattggc attgcatag ttgtatccat atcataatat gtacattat  
 3781 attggctcat gtccaacatt accgcatgt tgacattgat tattgactag ttataatag  
 3841 taatcaatta cggggtcatt agttcatag ccataatatg agttccgct tacataact  
 3901 acggtaaatg gccgccttg ctgacggccc aacgacccc gccattgac gtcaataatg  
 3961 acgtatgttc ccataagc gccaatagg actttccatt gacgtcaatg ggtggagat  
 4021 ttacggtaaa ctgcccactt ggcagtacat caagtgtatc atatgtcaag tacgcccc  
 4081 attgacgtca atgacggtaa atggccgccc tggcattatg cccagtacat gaccttatg  
 4141 ttcttcccta cttggcagta catctcagta ttagtcatg ttagtacct ggtgatg  
 4201 ttttggcagt acatcaatg gcgtggatag cggtttgaat caggggatt tccaagtct  
 4261 caocccattg acgtcaatg gagtttgttt tggcaccaaa atcaacggga ctttccaaa  
 4321 tgcgttaaca actccgccc attgacgcaa atggcggtg ggcgtgtac gtggagggtc  
 4381 tatataagca gagctcgttt agtgaacgt cagatcgct ggagacgcca tccacgctg  
 4441 tttagacctc atagaagaca ccgggaccga tccagctcc cggccggga acggtgcat

5

10

15

20

25

30

35

4501 ggaacgcgga ttccccgtgc caagagtgc gtaagtacgc cctatagact ctataggcac  
 4561 accccttgg cttctatgca tgctatactg tttttgctt gggccctata caccocgct  
 4621 tctttatgct ataggtagt gatatagctta gctataggt gtgggttatt gaccattatt  
 4681 gaccactccc ctattggtga cgatactttc cattactaat ccataacatg gctctttgcc  
 4741 acaactatct ctatggcta tatgccaata ctctgtcctt cagagactga caggactct  
 4801 gtatttttac aggatggggt cccatttatt atttacaat tccatatat acaacgcg  
 4861 tccccgtgc ccgcagtttt tattaaacat agcgtgggat ctccacgga atctcgggta  
 4921 cgtgttcggg acatgggctc ttctcggta gggcgggagc ttccacatcc gagccctggg  
 4981 cccatgcttc cagcggctca tggctgctcg gcagctcctt gctcctaaca gtggaggcca  
 5041 gacttaggca cagcaaatg cccaccacca ccagtgtgccc gcacaaggcc gtggcggtag  
 5101 ggtatgtgct tgaatatgag cgtggagatt gggctgcac gctgacgca gatggaagac  
 5161 ttaaggcagc ggcagaagaa gatgcaggca gctgagttgt tgtattctga taagagtca  
 5221 agttaactcc cgttcgggtg ctgttaacgg tggagggcag tgtagctcga gcagtactcg  
 5281 ttgctgcgc ggcgcacc agacataata gctgacagac taacagactg ttctttcca  
 5341 tgggtctttt ctgcagtcac cgtcggatca atcatcctc tegtgacttc ttctgtgtg  
 5401 gtgtttacct atatatctaa atttaattt tegtattata aaatttaata tatttcgacg  
 5461 atgaatttct caagatat tttctcgtg ttctgttgg ttctggcttt gtcaacagtt  
 5521 tcggctggc cagagccgaa aggtacccag gtgcagctgc aggagtcggg gggaggcttg  
 5581 gtaagccggg ggggtctcct tagagtctcc tgtgcagcct ctggattcac ttccagaaac  
 5641 gctggatga gctgggtccg ccaggctcca gggaaagggc tggagtgggt cggccgtatt  
 5701 aaagcaaaa ttgatgtggtg gacaacagac tatgctgcac ccgtgaaagg cagattccac  
 5761 atctcaagag atgattcaaa aaacacgtta tatctgcaa tgaatagcct gaaagccgag  
 5821 gacacagcgt tatattactg taccacgggg attatgataa ctttggggg agttatccct  
 5881 ccccgcaatt ggggccaggg aacctgggt accgtctcct cagctccac caagggccca  
 5941 tcgggtcttc cctgggcacc ctctccaag agcacctctg ggggcacagc ggccc:gggc  
 6001 accagcggcg aggaacttt tccgggtgtc ctacagtcct caggactcta ctctctagc  
 6061 aacgtggtga cgtgcccctc cagcagcttg ggcacccaga aagttgagc ccaatcttg tgacaaact  
 6181 cacaagccca gcaacacca ggtggacaag aaagttagc cctacatctg caacgtgaat  
 6241 cacacatgcc caccgtgccc agcacctgaa ctctggggg gaccgtcagt ctctctctc  
 6301 ccccccaaac ccaaggacac cctcatgac tcccgacc cttgaggtcac atgctgggtg  
 6361 gtggagctga gccacgaaga cctgaggtc agttcaact ggtacgtgga cggcgtggag  
 6421 gtgcataatg ccaagacaaa gccgcggggag gagcagtaca acagcagta ccgtgtgggtc  
 6481 agcctcctca cgtctctgca ccaggactgg ctgaatggca aggagtacaa gtgcaagggtc  
 6541 tcccaaaaag cctccacgc cccatcgag aaacctatc ccaagccaa agggcagccc  
 6601 cgagaacac aggtgtacac cctgccccca tcccgggatg agctgaccaa gaaccaggtc  
 6661 agcctgacct gctgggtcaa aggtcttat cccagcgaca tgcctgtgga gtggagagc  
 6721 aatgggcagc cggagaacaa ctacaagacc acgctctccg tegtggactc cgacggctcc  
 6781 ttcttctctc acagcaagct caccgtggac aagagcaggt ggcagcaggg gaacgtcttc

5

10

15

20

25

30

35

6841 tcatgtctcg tgatgcatga ggtctgtcac aaccactaca cgagaagag cctctccctg  
 6901 tctcgggta aagcgcaga gccgaaaaag cttctctatg agctgacaca gccacccctg  
 6961 gtgtcagtg cccagagaca aacggccagg atcacctgct ctggagatgc attgccagaa  
 7021 aaatatgttt attgtacca gcagaagtca ggcaggcccc ctgtggtggc catctatgag  
 7081 gacagaaac gaccctccg gatccctgag agattctctg gctccagctc agggacaatg  
 7141 gccacctga ctatcagtg gcccaggtg gaagatgaag gtgactacta ctgttactca  
 7201 actgcagca gtggttatca tagggaggtg ttaagcggag ggaccaagct gaccgtccta  
 7261 ggtcagcca aggtgcccc ctgggtcact ctgttcccac cctcctctga ggagcttcaa  
 7321 gccacaagg ccacactggt gtgtctcata agtgactcct accggggagc cgtgacagtg  
 7381 gcctggaag cagatagcag ccccgtcag gcgggagtgg agaccaccac accctccaaa  
 7441 caaagcaaca acaagtacgc ggcagcagc taccgtgacc tgacgcttga gcagtggaaag  
 7501 tcccacaaa gctacagctg ccaggtcacg catgaaggga gcacgttga gaagacagtg  
 7561 gcccctgcag aatgttcacc gcggaggag ggaaggcccc tttttgaagg gggaggaaac  
 7621 ttccggccat gactcctctc gtgcccccg caccggaacac tgatgtgcag agggccctct  
 7681 gccattgctg ctctcctctc ccttctcgt cactccgaat gtggcttctt tgctactgcc  
 7741 acagacaaga ataaaatctc aacatctaaa tgggttctct gagatttttc aagatcgctt  
 7801 aagcacattc ctctccagc accccttgct gcaggccagt gccaggcacc aacttggcta  
 7861 ctgctgcccc tgagagaaat ccagttcaat attttccaaa gcaaatgga ttacatatgc  
 7921 cctagatcct gattaacagg tgttttgtat tatctgtgct ttgcttccac ccacattatc  
 7981 ccattgctc cctcgaggg ggggcccggc acccaattcg cctatagtg agtcgtatta  
 8041 cgcgcgctca ctggcgtctg ttttacaag tegtgtactgg gaaaacccctg gcgttaccac  
 8101 acttaatcgc cttgcagcac atcccccttt cgcagctggt cgtaatagcg aagaggcccc  
 8161 caccgatcgc cttcccaac agttgcgag cttgaatggc gaatggaaat tgaagcgtt  
 8221 aatattttgt taaaattcgc gttaaaattt tgttaaatca gctcattttt taaccaatag  
 8281 gcgaaatcg gcaaatccc ttataaatca aagaataga ccgagatagg gttgagtgtt  
 8341 gtccagttt ggaacaagag tccactatta aagaacgtgg actccaaagt caaaggcgga  
 8401 aaaaccgtct atcaggcgga tggccacta ctccgggac atagacaag atgtgtatcc  
 8461 acctaaact aatgattttt accaaaatca ttaggggatt catcagtgtc cagggtcaac  
 8521 gagaattaac attcgtcag gaaagcttat gatgatgatg tgettataaa cttactcaat  
 8581 ggctgggttat gcatacgca atacatgca aaaacctaaa agagcttgcc gataaaaaag  
 8641 gccaatatt tgcatttac cgcggctttt tatgtagctt gaagataaa taaatagat  
 8701 aggttttatt tgaagtaaa tcttcttat cgtaaaaaat gccctcttgg gttatcaaga  
 8761 gggctattat atttcgagg ataacatcat ttggtgacga ataaactaag cactgtctc  
 8821 ctgtttactc cctgagctt gaggggttaa catgaaggtc atcgatagca ggataataat  
 8881 acagtataac gctaaaccaa taatccaat ccagccatcc caaattgga ttgaatgatt  
 8941 ataataaca gcaaacagta atgggccaat aacaccggtt gcattggtaa ggtcaccaa  
 9001 taatccctgt aagcacctt gctgatgact cttgttttgg atagacatca ctccctgtaa  
 9061 tgcaggtaaa gcgatccac caccagccaa taaaattaaa acaggggaaa ctaaccaacc  
 9121 ttccagatata aacgctaaaa aggcataatgc actactatct gcaataatc cgagcagtac

5

10

15

20

25

30

35

9181 tgcggttttt tgcgccatlt agtggctatt ctctctgcca caaaggcttg gaatactgag  
 9241 tgtaaaagac caagaccgtt aatgaaaagc caaccatcat gctattcatc atcagatttt  
 9301 ctgttaatagc accacaccgt gctggatggg ctatcaatgc gctgaaataa taatacaaca  
 9361 atggcatcgt taataagtg atgtataccg atcagctttt gtcccttta gtgagggtta  
 9421 ttgtgcgctt tggcgtatc atgggtcatag cgtttctctg tgtgaaatgg ttatccgctc  
 9481 acaattccac acaacatacg agccggaagc ataaagtgtg agccttgggg tgcctaataga  
 9541 gtgagtaaac tcacattaat tgcgttgcgc tcaactgccc ctttccagtc gggaaacctg  
 9601 tctgtccagc tgcattaatg aatcggccaa cgcgcgggga gaggcgggtt cgtatttggg  
 9661 cgtctttccg ctctctcgt cactgactcg ctgcgtcggg tctgttcggc ggcgcgagcg  
 9721 gtatcagctc actcaaaagg ggttaatacgg ttatccacag aatcagggga taacgcagga  
 9781 aagaacatgt gagcaaaaagg ccagcaaaaag gccaggaaacc gtaaaaaggc cgcgttgctg  
 9841 gcgtttttcc ataggctccg cccccctgac gagcatcaca aaaatcgagc ctcaagtcag  
 9901 aggtggcgaa accgcacagg actataaaga taccaggcgt ttccccctgg aagctccctc  
 9961 gtgcgctctc ctgttccgac cctgcgctt accgga\*acc tgtccgctt tctcccttcc  
 10021 ggaagcgttg cgttttctca tagctcacgc cccgttcagc cgcaccgctg cgccttatcc  
 10081 cgttccaaagc tgggtgtgtg gcacgaaccc cccgttcagc cgcaccgctg cgccttatcc  
 10141 ggttaactatc gtcttgagtc caaccggga agacacgact tatcgccact ggcagcagcc  
 10201 actggtaaca ggtatagcag agcagggtat gtaggcgggtg ctacagagtt cttgaagtgg  
 10261 tggcctaact accgtacac tagaaggaca gtatttggta tctgcgctct gctgaagcca  
 10321 gtacacttcc gaaaagagt tggtagctct tgatccggca acaaaaccac cgttggtagc  
 10381 gttgggtttt ttgttgcaa gcagcagatt acgcgagaa aaaaaggatc tcaagaagat  
 10441 cttgtgatct tttctacggg gtctgacgct cagtgaacgc aaactcacg ttaagggtatt  
 10501 tttaaatcaa tctaaagtat atatgagtaa acttggctct acagttaaca atgcttaatc  
 10561 atgagggcac ctatctcagc gatctgtcta ttctgttcat ccatagttcg ctgacacccc  
 10621 gtctgttaga taactacgat acgggaggggc ttaccatctg gccccagtgc tgaatgata  
 10741 ccgcgagacc cagctcacc ggtccagat ttatcagcaa taaaccagcc agccggaagg  
 10801 gccgagcgca gaagtgttc tgcacttta tccgcctcca tccagctcat taattgttcg  
 10861 cgggaagcta ggttaagtag ttcgcaggtt aatagtttg gcaacgttgt tgcatttgtt  
 10921 acaggcatcg tgggtgtcacg ctctcgtttt ggtatggctt cattcagctc cggttcccaa  
 10981 cgatcaaggc ggtttacatg atcccccatg ttgtgcaaaa aagcgtttag ctccctcggg  
 11041 cctccgatcg ttgtcagaag taagttggcc cagtggttat cactcatggt tatggcagca  
 11101 ctgcataatt ctcttactgt catgccatcc gtagctatc ttctgtgac tgggtgagtag  
 11161 tcaaccaagt cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgtca  
 11221 atacgggata ataccgccc acatagcaga acttcaaaag tgcctcatcat tggaaaacgt  
 11281 tcttcggggc gaaaactctc aagatcttta cgcgtgttga gatccagttc gatgtaaccc  
 11341 actcgtgcac ccaactgac ttacagcatct ttacttttca ccagcgtttc tgggtgagca  
 11401 aaaaaggaa ggcataatgc cgcaaaaaag ggaataaagg cgacacggaa atgttgaata  
 11461 ctcatactct tcttttttca atattattga agcatttatc aggggttatg tctcatgagc

5

10

15

20

25

30

35

11521 ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc  
 11581 cgaaaagtgc cac

SEQ ID NO:102 (pTnMod (CHOvep-prepro-HCPro-CPA))

5           1   ctgagcgcc ctgtagcgcg gattaaagc cgcggggtgt ggtggttac cgcagcgtga  
           61   ccgctacat tgcagcgcc ctgagcgccg ctcttctgc ttcttccct tcccttctcg  
           121   ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgatc catatcataa  
           181   tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac  
           241   tagttattaa tagtaataa ttacggggtc attagtcat agccatata tggagttccg  
           301   cgttacataa cttacggtaa atggccgcgc tggctgaccg ccaacgacc cccgccatt  
           361   gacgtcaata atgacgtatg ttcccatagt aacccaata gggactttcc attgacgtca  
           421   atgggtggag tatttacggt aaactgccc cttggcagta catcaagtgt atcatatgcc  
           481   aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta  
           541   catgacctta tgggactttc ctacttgga gtacatrtac gtattagta tgcctattac  
           601   catggcgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcagggg  
           661   atttccaagt ctccaccca ttgacgtcaa tgggagtgtg tttggcacc aaatcaacg  
           721   ggactttcca aaatgtcga acaactcgc cccattgacg caaatggcg gtaggggtgt  
           781   acgggtggag gtctatataa gcagagctcg tttagtgaa cgtcagatcg cctggagacg  
           841   ccatccacgc tgttttgacc tccatagaag acaccgggac cgtaccagcc tccgcgccg  
           901   ggaacggtgc attggaacgc ggaattcccc tgccaagagt gacgtaagta ccgctatag  
           961   actctatagg cacacccct tggctcttat gcatgctata ctgttttttg cttggggcct  
           1021   ataaccccc gcttctctat gctataggtg atggtatagc ttagcctata ggtgtgggtt  
           1081   attgacctt attgacct cccctattgg tgacgatact ttccattact aatccataac  
           1141   atggctcttt gccacaacta tctctattgg ctatatgcca atactctgc ctccagagac  
           1201   tgacacggac tctgtatatt tacaggatgg ggtcccatat attattaca aattcacata  
           1261   tacaacaacg ccgtcccccg tgcccgcagt tttattaaa catagcgtgg gatctccacg  
           1321   cgaatctcgc gtacgtgttc cggacatggg ctcttctccg ttagcggcgg agcttccaca  
           1381   tccagcccc ggtcccatgc ctccagcggc tcatggctgc tggcagctc cttgctccta  
           1441   acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag  
           1501   gccgtggcgg taggggtatgt gtctgaaat gagcgtggag attgggctcg cacggctgac  
           1561   gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgatttc  
           1621   tgataagagt cagaggtaac tcccgcttgc gtgctgttaa cgggtggagg cagtgtagtc  
           1681   tgacagtagt tctgtctgc cgcgcgcgc accagacata atagctgaca gactaacaga  
           1741   ctgttccctt ccatgggtct tttctgcagt caccgtcga ccatgtgtga acttgatatt  
           1801   ttacatgatt ctctttacca attctgccc gaattacact taaacgact caacagctta  
           1861   acgttggctt gccacgcatl acttgactgt aaaactctca ctctaccga acttggccgt  
           1921   aacctgccaa ccaagcgag aacaaaacat aacatcaaac gaatcgacgg attgttaggt  
           1981   aatcgtcacc tccacaaaga ggcactcgct gtataccgtt ggcagctag ctttatctgt  
           2041   tcgggcaata cgatgccat tgtacttgtt gactggctcg atattctga gcaaaaacga

2101 cttatgggtat tggagagcttc agtcgcacta cagggctcgt ctgttactct ttatgagaaa  
 2161 gggttcccg tttcagagca atgttcaaaag aaagctcatg accaatctct agccgacctt  
 2221 gcgagcattc taccgagtaa caccacacog ctcattgtca gtgatgctgg ctttaaagtg  
 2281 ccattgtata aatccgttga gaagctgggt tggfactggg taagtcgagt aagaggaaaa  
 2341 gtacaatatg cagaccctagg agcggaatac tggaaacctc tcagcaactt acatgatatg  
 2401 tcatctagtc actcaagagc tttagggctat agagggctga ctaaaagcaa tccaatctca  
 2461 tggcaaatc tattgtataa atctcgtctt aaagggccgaa aaatcagcg ctgcacacgg  
 2521 actcattgtc accaccgtc acctaaaatc tactcagcgt cggcaaaagg gccatggggtt  
 2581 ctageaacta acttacctgt tgaatttcga acaccaaac aactgtttaa tatctattcg  
 2641 aagcgaatgc agattgaaga aaccttcga gacttgaaaa gtccgtccta cggactaggc  
 2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcattgctgt aatcgccctg  
 2761 atgcttcaac taactgtgtg gcttgcgggc gtctcatgctc agaaacaagg ttgggacaag  
 2821 cacttccagg ctaacacagt cagaaatcga aacttactct caacagtctg cttagggcacg  
 2881 gaagtcttgc ggcattcttg ctacacaata acaaggaaag acttactcgt ggtggcaacc  
 2941 ctactagctc aaattttatt cacacatggt tacgctttgg ggaattatg ataatgatcc  
 3001 agataccttc tggctaataa aagatcagag ctctagagat ctgtgtgttg gtttttttg  
 3061 gatctgctgt gccttctagt tggcagccat ctgtgtgttg cccctcccc gtgcttctct  
 3121 tgaccctgga aggtgcact cccactgtcc ttccctaata aaatgaggaa attgcatcgc  
 3181 atgtcttgag taggtgtcat tctattcttg ggggtgggggt ggggcagcac agcaaggggg  
 3241 aggatctgga agacaatagc aggcattgctg gggatgggggt gggctctatg ggtacctctc  
 3301 tctctctctc tctctctctc tctctctctc tctctctctc tctctctctc tctctctctc  
 3361 tctctctctc tctctctctc tgggtaccag gtgctgaga attgacccgg tgacaaaagg  
 3421 tgccttttat cactactta aaataaaaa acaattactc agtgcctgtt ataagcagca  
 3481 attaatatg attgatgcct acatcaaac aaaaactgat ttaacaaatg gttgggtctgc  
 3541 cttagaaaag atattgaac attatcttga ttatatattt gataataata aaacutttat  
 3601 cccatcccaa gaagtgtgc ctatcattgg ttggaatgaa ctgaaaaaaa attagccttg  
 3661 aatacatcac tggtaaggta aacgaccttg tcagcaaat gatccaagag aaccaactta  
 3721 aagcttctct gacggaatgt taattctcgt tgacctgag cactgatgaa tccctaatg  
 3781 atttttgtaa aaatcattaa gtaagggttg atacacatct tgtcatatga tcccggtaat  
 3841 gtgagtttag tcaactatta ggcaccccg gctttacact ttatgcttcc ggtcgtatg  
 3901 ttgtgtgtaa ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac  
 3961 gcaagcgcg caattaacc tcaataaagg gaacaaaagc tggagctcca ccgctgtggc  
 4021 ggcgctcta gaactagtgg atccccggg ctgcaggaaat tcatatcaa gottatogat  
 4081 accgctgacc tcgagctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag  
 4141 ctgacactga taactgattt tctcaaaa cgggaacaa cacaatccca caaacagct  
 4201 cagagagaaa caatcactga tggctacagc accaagggtat gcaatggcaa tccattogac  
 4261 attcatctgt gacctgagca aaatgatttc tcttccatg aatggttctt tcttccctc  
 4321 atgaaaaagg aatttccaca ctcaaatat gcgacaaaga caaacagaga acaattaatg  
 4381 tgcctcttcc taatgtcaaa attgtagtgg caaaggagg acaaaaatct caagttctga

5

10

15

20

25

30

35

4441 gtagggtttta gtgattggat aagaggcttt gacctgtgag ctcacctgga cttcatatcc  
 4501 ttttggataa aaagtgcctt tataactttc aggtctccga gtctttattc atgagactgt  
 4561 tgggttttaggg acagaccac aatgaaatgc ctggcatagg aaagggcagc agagccttag  
 4621 ctgacctttt ctgtggacaa gcatgttcaa caatgtgtg aaaaactat ttgtactgct  
 4681 ttgcacagct gtgtgggca gggcgatcca ttgccacctc tcccaggtaa ccttcaact  
 4741 gcaagaagat tgtgtgttac tctctctaga agcttctgc agactgacat gcatttcata  
 4801 ggtagagata acatttactg ggaagcacat ctatcatcac aaaaagcagg caagattttc  
 4861 agactttctt agtggctgaa atagaagcaa agacgtaat taaaacaaa atgaaacaaa  
 4921 aaaaatcagt tgatactgtt ggtgtagaca ccagcaaaa aaatatattt tgcactacca  
 4981 tcttgtctta agtctcaga cttagcaagg agaattgtaga ttccacagt atatatgttt  
 5041 tcacaaaagg aaggagagaa acaaaagaaa atggcactga ctaaacttca gctagtggta  
 5101 taggaagta attctgttta acagagatgt cagtgtatc tatgtatgtc ctgaagaatt  
 5161 atgtgttact tttttcccc attttttaat caaacagtgc ttacacaggg tcagaatgg  
 5221 tcttctactg tttgtcaatt ctattatttc aatacaaac aatagcttct ataactgaaa  
 5281 tatatttgc attgtatatt atgattgtcc ctogaacct gaacactcct ccagctgaat  
 5341 ttcacaattc ctctgtatc tggcaggcca ttaagtatt catggaagat ctttgaggaa  
 5401 cactgcaagt tcatatcata aacacatttg aaattgagta ttggtttgca ttgtatggag  
 5461 ctatgttttg ctgtatctc agaaaaaag ttgttataa agcatttaca ccataaaaa  
 5521 gatagattta aatatccag ctataggaaa gaaagtgcgt ctgctcttca ctctagtctc  
 5581 agtgggtctc ttcacatgca tgccttctta ttctctctat ttgtcaaga aataatagg  
 5641 tcaagctctg tctcaacta tgcctctgct agcatggctc agatgcacgt tgtagataca  
 5701 agaaggatca aatgaacag acttctggtc tgtacctac aacctatgta ataagcacac  
 5761 taactaataa ttgctaatta tgttttccat ctctaagggt cccatatatt tctgttttct  
 5821 taaagatccc attatctggg tgaactgaa gctcaatgga acatgagcaa tatctccag  
 5881 tcttctctcc catccaacag tctgtatgga ttgacgaac aggcagaaaa cacattgtta  
 5941 cccagaatta aaactaata ttgtctctcc attcaatcca aaatggacct ttgaaaacta  
 6001 aaatctaacc caatccatt aaatgatttc tatggtgtca aaggtcaaac ttctgaaggg  
 6061 aacctgtggg tgggtcaciaa ttcaggctat atattcccca gggctcagcc agtggatcaa  
 6121 tcatctatct cgtgacttct tctgtgtggt tgtttacctc tatatctaaa ttaatatatt  
 6181 cgtttattaa aatttaatat atttcgacga tgaatttctc agggatatatt ttcttcgtgt  
 6241 tctgttttgg tctggttttg tcaacagttt cggctgcgcc agagccgaaa ggtaccacag  
 6301 tgaagctgca ggaagtcggg ggaagcttgg taaagccggg ggggtccctt agagtctcct  
 6361 gtgcagcctc tggattcact ttacgaacg cctggatgag ctgggtccgc caggctccag  
 6421 ggaaggggct ggaaggggtc ggcggtatta aagcaaaaat tgaatgtggg acaacagact  
 6481 atgctgcacc cgtgaaggc agattcacca tctcaagaga tgaattcaaaa aacacgttat  
 6541 atctgcaaat gaatagcctg aaagccgagg acacagcgt atattactgt accacgggga  
 6601 ttaatgataac atttggggga gttatccctc cccgaattg gggccaggga accctgggtca  
 6661 cgtctctctc agcctccacc aaggggccat cggctctccc cctggcaccc tccccaaga  
 6721 gacctctggt gggcacagcg gccctgggct gcctgggtcaa ggactacttc ccgaacagg

5

10

15

20

25

30

35



6781 tgcggtgtgc gtggaactca ggggcccctga ccagggggcgt gcacaccttt ccggctgttcc  
6841 tacagctctc aggaactctac ttcccttagca acgtgggtgac cgtgccctcc agcagcttgg  
6901 gcaccagac ctacatctgc aacgtgaatc aacgtcccag caacaccaag gtggacaaga  
6961 aagttgagcc caaatcttgt gacaaaactc acacatgccc acgtgccc gcacctgaac  
7021 tccctggggg accgtcagtc ttccctcttc ccccaaac caaggacac ctcatgatct  
7081 cccggacccc tgaggtcaca tgcgtggtgg tggacgtgag ccacgaagac cctgaggtca  
7141 agttcaactg gtacgtggac ggcgtggagg tgcataatgc caagacaaag ccgcgggagg  
7201 agcagtacaa cagcacgtac cgtgtgttca ggtctctcac cgtctgac caggactggc  
7261 tgaatggcaa ggaatcaag tgcaaggtct ccaacaaagc cctcccagc ccatcgaga  
7321 aaacatctc caagccaaa gggcagccc gagaaccaca ggtgtacac ctgccccat  
7381 cccgggatga gctgaccaag aaccaggtca gctgacctg cctgggtcaa ggtctctatc  
7441 ccagcgacat cgccgtggag tgggagagca atgggcagcc ggagaacaaac tacaagacca  
7501 cgctccctg gctggactcc gacggctcct tcttctctca cagcaagctc accgtggaca  
7561 agacaggtg gcagcagggg aacgtcttct catgctcgt gatgcatgag gctctgcaca  
7621 accactacac gcagaagagc ctctccctgt ctccggttaa agcgcagag ccgaagcttt  
7681 cctatgagct gacacagcca ccctcgggtg cagtgtcccc aggacaaaag gccaggatca  
7741 cctgctctgg agatgcatcg ccagaaaaat atgtttattg gtaccagcag aagtcaggcc  
7801 aggcctctgt ggtgtctatc tatgaggaca gcaaacgacc ctccgggac cctgagagat  
7861 tctctggctc cagctcaggg acaatggcca ccttgactat cagtggggcc caggtggaag  
7921 atgaaggtga ctactactgt tactcaactg acagcagtg ttatcatagg gaggtgttca  
7981 ggggagggg caagctgacc gtccataggtg agcccaaggc tgcctctcg gtcactctgt  
8041 tccacacctc ctctgaggag cttaagcca acaaggccac actgggtgtg ctcataagtg  
8101 actcctaccc gggagccgtg acagtggcct ggaaggcaga tagcagcccc gtcaaggcgg  
8161 gagtggagac caccacacc tccaaacaaa gcaacacaa gtacggggc agcagctacc  
8221 tgagcctgac gcttgagcag tggagtgccc ctgcagaatg ttaccgcgg agggagggaa  
8281 aaggagacac cgtggagaag acagtggccc cgtcagaatg ttaccgcgg agggagggaa  
8341 gggccctttt tgaaggggga gaaactctg cgcatgact cctctcgtg ccccgacg  
8401 gaacactgat gtgcagaggg cctctgcca ttgctgttc ctctgacct cctcgtcact  
8461 ctgaatgtgg ctctcttctg actgccacag caagaataa aatctcaaca tctaaatggg  
8521 ttctctgaga tttttcaaga gtggttaagc acatctctc cccagacccc cttgctgcag  
8581 gccagtcca ggcaccaact tggctactgc tgccatgag agaaatccag ttcaatatt  
8641 tccaaagcaa aatggattac atatgcccta gatcctgatt aacaggtgtt ttgtattatc  
8701 tgtgtcttcg cttaccacc attatcccat tgcctcccc cagggggggg cccggtaccc  
8761 aactgcccct atagttagtc gtattacgag cgctcactgg ccgtcgtttt aacagtcgt  
8821 gactgggaaa accctggcgt taccacactt aatgccttg cagcacatcc cctttcgc  
8881 agctggcgta atagcgaaga gggccgacc gatgcctt ccaacagtt gcgcagcctg  
8941 aatggcgaat ggaaattgta agcgttaata tttgtttaa attcggtta aattttgtt  
9001 aaatcagctc atttttaa caataggcg aaatcggcaa aatccctat aaatcaaaag  
9061 aatagaccga gatagggttg agtgtgttcc cagtttgga caagagtcca ctattaaaga

5

10

15

20

25

30

35

9121 acgtggactc caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactactcc  
 9181 gggatcatat gacaagatgt gtatccacot taacttaatg atttttacca aaatcattag  
 9241 gggatcattc agtgcacagg gtcaacgaga attaacattc cgtcaggaaa gcttatgatg  
 9301 atgattgtgt taaaaactta ctcaatggct ggttatgcat atcgcaatc atcgcaaaaa  
 9361 cctaaaagag cttgcgata aaaaaggcca atttattgct atttaccgag gctttttatt  
 9421 gagcttgaaa gataaataaa atagatagggt ttattttgaa gctaaatctt ctttatcgta  
 9481 aaaaatgccc tcttggttta tcaagaggggt cattatatct cgcggaataa catcatttgg  
 9541 tgacgaaata actaagcact tgcctcctgt ttaactcccc gagcttgagg ggttaacatg  
 9601 aaggtcatcg atagcaggat aataatcacg taaaacgcta acccaataat ccaatccag  
 9661 ccatcccaa ttggtagtga atgattataa ataacagcaa acagtaatgg gccataaaca  
 9721 ccggttgcat tggtaaggct caccaataat cctgttaaag caccttgctg atgactcttt  
 9781 gtttgatag acatcactcc ctgtaatgca ggtaaagcga tcccaccacc agccaataaa  
 9841 attaaaaacag gaaaaactaa ccaaccitca gatataaaacg ctaaaaaggc aaatgcacta  
 9901 ctatctgcaa taaatccgag cagtactgcc gttttt\*ogc ccatttagtg gctattcttc  
 9961 ctgccacaaa ggcttggaaat actgagtgtg aaagaccaag accgtaatg aaagccaac  
 10021 caatcgcta ttcacatca cgatttctgt aatagacca caccgtgctg gatggctat  
 10081 caatcgctgt aaataataat caacaaatgg catcgttaaa taagtgtgtg ataccgatca  
 10141 gcttttgttc ccttagtga gggttaatgg cgcgcttgcc gtaatcatgg tcatagctgt  
 10201 ttcctgtgtg aaatgcttat ccgctcaca ttccacacaa catcagagcc ggaagcataa  
 10261 agtgaataag ctgggtgccc taatgagtga gctaaactc acatattggg ttggtgtcac  
 10321 tgcgcgcttc ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaaacgag  
 10381 cggggagagg cgggttgctg attggcgct cttcggcttc ctgctcact gactcgtgc  
 10441 gctcggctgt tcggtgcgg cagcgtggtat cagctcactc aaaggcgtta atacggttat  
 10501 ccacagaatc aggggataac gcaggaaaga acatgtgagc aaaaaggccag caaaagcca  
 10561 ggaaccgtaa aaaggccgag ttgctggcgt ttttccatag gctccgccc cctgacgagc  
 10621 atcacaaaaa tgcagcctca agtcagaggt ggcgaaccc gacaggacta taaagatacc  
 10681 aggcgtttcc ccttggaagc tccctcgtgc gctctcctgt tccgacctg ccgcttacg  
 10741 gatacctgtc cgcctttctc ccttgggaa gcgtggcgct ttctcatagc tcaagctgta  
 10801 ggtatctcag ttccgtgtag gtctgtcgt ccaagctggg ctgtgtgcac gaacccccg  
 10861 ttcagccoga ccgtgcgccc ttatccggtt actatgctct tgagtccaac ccggttaagac  
 10921 acgacttacc gccactgga gcagccactg gtaacaggat tagcagagcg aggtatgtag  
 10981 aggttcttac agatctcttg aagtgtggc ctaactacgg ctacactaga agacagtat  
 11041 ttggtatctg cgtctgtgct agccagttta ccttcgaaa aagagtgggt agctcttgat  
 11101 ccggcaaaaa aaccaccgct ggtagcgggt gttttttgt ttgcaagcag cagattacgc  
 11161 gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacgggtctt gacgtcagt  
 11221 ggaacgaaaa ctaacgttaa gggattttgg tcatggatt atcaaaaaag atcttcacct  
 11281 agatcctttt aaattaaaaa tgaagtttta aatcaacta agtatatat gagtaaaactt  
 11341 ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctacagcgtc tgcattttc  
 11401 gttcatccat agtgcctga ctcccgctg ttagataaac tacgatacgg gagggcttac

11461 catctggccc cagtgtgtgca atgataccgc gagacccaacg ctacccgggt ccagatttat  
 11521 cagcaataaa ccagccagcc ggaaggcccg agcgcagaag tggctctgca actttatccg  
 11581 cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttaata  
 11641 gtttgccgaa cgttgttgc attgtacag caatcgtggt gtcacgctcg tcggtttggtta  
 11701 tgggttcatt cagctccggt tcccaacgat ccagtcgagt tacatgatcc cccatgttgt  
 11761 gcaaaaaagc ggttagctcc ttcggctctc cgatcgttgt cagaagtaag ttggccgcag  
 11821 tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg ccatacgtaa  
 11881 gatgttttc tgtgactggt gagtactcaa ccaagtcatt ctgagaatag tgtatgccc  
 11941 gaccagttg ctcttgccc ggtcaatac gggataatac cggccacat agcagaactt  
 12001 taaaagtgt catcattgga aaacgttctt cggggcgaaa actctcaagg atctaccgc  
 12061 tgttgagatc cagttcgatg taaccactc gtgcaccaa ctgatcttca gcattcttta  
 12121 ctttcaccag cgtttctggg tgagcaaaa caggaaggca aaatgccgca aaaaagggaa  
 12181 taaggcgac acggaatgt tgaatactca tactcttctt tttcaatat tatgaaagca  
 12241 tttatcaggg ttatgtctc atgagcggat acatat.tga atgtatttag aaaataaac  
 12301 aaataggggg tccgcgcaca ttccccgaa aagtgccac

SEQ ID NO:103 (pTnMod (CHOrep-prepro-HCPro-LYS-CPA))

1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtgggttacg cgcagcgtga  
 61 ccgtacact tgccagcgc ctgagcccg ctcttttcgc tttcttccct tcttttctcg  
 121 caactgtgc cggcatcaga ttggtctattg gccattgcat agttgtatc catatcataa  
 181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac  
 241 tagttattaa tagtaataca ttacggggtc attagttcat agccatata tggagtccg  
 301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc ccgcccatt  
 361 gcgtcaata atgacgtatg ttccatagt aacgccaata gggactttcc attgacgtca  
 421 atgggtggag tattacggt aaactgcca cttggcagta catcaagtgt atcatatgac  
 481 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta  
 541 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tegtattac  
 601 catggtgatg cggttttggc agtacatcaa tggcggtgga tagcgttttg actcacgggg  
 661 atttccaagt ctccaccca ttgacgtcaa tgggagtgtg ttttggcacc aaatcaacg  
 721 ggactttcca aaatgtcgt acaactccgc ccaattgacg caaatggcg gtaggcgtgt  
 781 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg  
 841 ccatccacgc tgttttgacc tccatagaag acacggggac cgatccagcc tccgcggccg  
 901 ggaacggtgc attggaacgc ggattcccc tgccaagagt gacgtaagta ccgcctatag  
 961 actctatagg cacaccctt tggctcttat gcagtctata ctgttttttg cttggggcct  
 1021 atacacccc gcttctttat gctatagggt atggtatagc ttagectata ggtgtgggtt  
 1081 atggaccatt attgaccact cccctattgg tgaagatact ttccattact aatccataac  
 1141 atggctttt gccacaaacta tctctattgg ctatatgcca atactctgc cttcagagac  
 1201 tgacacggac tctgtatttt tacaggatgg ggtccattt attattaca aattcacata  
 1261 tacaacacg ccgtcccccg tgcgccgagt ttttattaaa catagcgttg gatctccacg

1321 cgaatctcgg gtacgtgttc cggacatggg ctctctctcg gtagcggcgg agcttccaca  
 1381 tccgagccct ggtcccatgc ctccagcggc tcatggctcg tcggcagctc cttgctccta  
 1441 aagtgaggg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag  
 1501 gccgtggcgg tagggtagt gtctgaaat gagcgtggag attgggctcg cacggctgac  
 1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgatctc  
 1621 tgataagagt cagaggtaac tcccgctggg gtgctgttaa cggctggagg cagtgtagtc  
 1681 tgagcagtrac tegtgtctgc cgcgcgcgc accagacata atagctgaca gactaacaga  
 1741 ctgttctctt ccatgggtct tttctgcagt caccgtcgga ccatgtgtga acttgatatt  
 1801 ttacatgatt ctctttacca attctgcccc gaattacact taaaagact caacagctta  
 1861 acgttggctt gccacgcatt acttgactgt aaactctca ctctaccga acttggcctg  
 1921 aacctgcaa ccaagcggag aaaaaacat aacatcaac gaatcgaccg attgttaggt  
 1981 aatcgtcacc tccacaaaga ggcactcgt gtataccgtt ggcagtctag ctttatctgt  
 2041 tcgggcaata cgatgccat tgtacttgtt gactggcttg atattcgtga gcaaaaacga  
 2101 cttatgggtat tgcgagcttc agtcgcacta cagggt~gtt ctgttactct ttatgagaaa  
 2161 gcgttccccc ttccagagca atgttcaag aaagcccatg accaatttct agcgcacctt  
 2221 gcgagcattc taccagtaga caccacaccy gctattgtca gtgatgtctg cttttaaagt  
 2281 ccatgggata aatcgttga gaagctgggt tggtaactgt taagtcgagt aagaggaaaa  
 2341 gtacatatg cagacctagg agcggaaaaa tggaaacctc tcagcaactt acatgatatg  
 2401 tcatctagc actcaagac tttaggttat aagaggtcga ctaaaagcaa tccaatctca  
 2461 tgcaaaatc tattgtataa atctcgtctt aaaggccgaa aaaatcagcg ctgcacacgg  
 2521 actcattgtc accaccgtc acctaaaac tactcagcgt cggcaaaagg gccatgggtt  
 2581 ctagaacta acttaactgt tgaattcga acaccacaa aacttgtaa tctctattcg  
 2641 aagcgaatgc agattgaaga aaccttccga gacttgaata gtctgacctc cggactaggc  
 2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgtctgt aatcgccctg  
 2761 atgcttcaac taacatgttg gcttgcgggc gtccatgctc agaaacaagg ttgggacaa  
 2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagtctg cttaggcattg  
 2881 gaagttttgc ggcattcttg ctacacaata acaagggaag acttactcgt ggcctgaacc  
 2941 ctactagctc aaaatttatt cacacatgggt tacgcttttg ggaattatg ataatgatcc  
 3001 agatcacctc tggctaataa aagatcagag ctctagagat ctgtgtgttg gttttttgtg  
 3061 gatctgctgt gccctctagt tggcagccat ctgttgtttg cccctcccc gtgcttctct  
 3121 tgaccttga aggtgccact cccactgtcc ttcttaata aaatgaggaa attgcatcgc  
 3181 attgtctgag tagtgtcat tctattcttg ggggtgggggt ggggagcac agcaaggggg  
 3241 aggatggga agacaatagc aggcagtctg gggatgcgggt gggctctatg ggtacctctc  
 3301 tctctctctc tctctctctc tctctctctc tctctctctc cctctctctc tctctctctc  
 3361 tctctctctc tctctctctc tctctctctc tctctctctc tctctctctc tctctctctc  
 3421 tgccttttat catcacttta aaaaataaaa acaattactc agtgcctgtt ataagcagca  
 3481 attaatatg attgatgect acatcacac aaaaactgat ttaacaaatg gttgggtctg  
 3541 cttagaaagt atatttgaac attatcttga ttatatatt gataataata aaaccttat  
 3601 ccctatccaa gaagtgtgc ctatcattgg ttggaatgaa cttgaaaaaa attagccttg

5

10

15

20

25

30

35

3661 aatacattac tggtaaggta aacgccattg tcagcaatt gatccaagag aaccaactta  
 3721 aagctttcct gacggaatgt taattctcgt tgaccctgag cactgatgaa tccctaatag  
 3781 attttggtaa aaatcattaa gtaagggtgg atacacatct tgtcatatga tcccggtaat  
 3841 gtggttagc tcaatcatta ggcacccag gctttacact ttatgcttcc ggctcgatg  
 3901 ttgtgtgaa ttgtgagcgg taacaattt cacacaggaa acagctatga ccatgattac  
 3961 gccaaagcgg gaattaacc tcactaaagg gacaaaagg tggagctcca ccgcggtggc  
 4021 ggcgctcta gaactagtgg atccccggg gcgcaggagt tgcataatga gcttatcgat  
 4081 accgtgacc tcgagctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag  
 4141 cttgacctga taccgattt tcttcaaca ggggaaacaa cacaatccca caaacacgct  
 4201 cagagagaaa ccatcactga tggctacagc accaaggat gcaatggeaa tccattcgac  
 4261 attcatctgt gacctgagca aaatgatttc tctctccatg aatgggtgct tcttccctc  
 4321 atgaaaaggc aatttccaca ctcacaatat ggcacaaaga caaacagaga acaattaatg  
 4381 tgcctcttcc taatgtcaaa attgtagtgg caaagaggag acaaaaatct caagtcttga  
 4441 gtagggttta gtgattggat aagaggcttt gacctg-gag ctcacctgga cttcatatcc  
 4501 ttgtggataa aaagtgcctt tataacttcc aggtctccga gcttttattc atgagactgt  
 4561 tgggttaggg acagaccac aatgaaatgc ctggcatagg aaagggcagc agagccttag  
 4621 ctgacctttt ctggggacaa gcatgtgcaa acaatgtgtg acaaaactat ttgtactgt  
 4681 ttgcacagct gtgctgggca gggcgatcca ttgccacta tccaggttaa ccttccaaat  
 4741 gaagaagat tgttgcctac tctctctaga aagctctgc agactgacat gatttccata  
 4801 ggtagagata acatttactg ggaagcacat ctatcatcac aaaaagcagg caagattttc  
 4861 agactttcct agtggctgaa atagaagcaa agagctaat taaaaacaaa atgaaacaaa  
 4921 aaaaactcagt tgataacctgt tgcagcaaaa aaatatattt tgcactacca  
 4981 tcttgcctta agtccctcaga cttagcaagg agaattgtaga ttccacagt atatatgttt  
 5041 tcacaaaagg aaggagagaa acaaaaagaa atggcactga ctaaaactca gtagtggtta  
 5101 taggaaagta attctgctta acagagattg cagtgatctc tatgtatgtc ctgaagaatt  
 5161 atgttgtact tttttccccc atttttaaat caaacagtgc ttacagagg tcagaatgggt  
 5221 tcttttactg ttgttcaatt ctattatttc aatacagaac aatagcttct ataactgaaa  
 5281 tatatttgc attgtatat attgattgtcc ctgcaacct gaacactcct ccagctgaat  
 5341 ttcacaattc ctctgtatc tgcagggcca ttaagttatt catggaagat ctttgaggaa  
 5401 cactgaagt tcatatcata aacacatttg aaattgagta ttggtttgca ttgtatggag  
 5461 ctatgttttg ctgtatctc agaaaaaaag ttgttataa agcattcaca cccataaaaa  
 5521 gatagattta aatattccag ctataggaaa gaaagtgcgt ctgctcttca ccttagtctc  
 5581 agttggctcc ttcacatgca tgccttctta tttctcctt ttgtcaaga aaataatagg  
 5641 tcaagcttgg tctcactta tgcctgcct agcatggctc agatgcacgt tgcagataga  
 5701 agaagatca aatgaaacag acttctggct tgttacctac accatagta ataagcac  
 5761 taactaataa ttgctaatta tgttttccat ctctaagggt cccatatttt tctgttttct  
 5821 taaagatccc attatctggt tgtaactgaa gctcaatgga acatgagcaa tatttccag  
 5881 tcttctctcc catcoaacag tctgtatgga ttgacagaac aggcagaaaa cacattgtta  
 5941 cccagaatta aaactaata ttgtctctcc attcaatcca, aaatggacct attgaaacta

5

10

15

20

25

30

35

6001 aaatctaacc caatccatt aatgatttc tatgtgttca aaggtcaaac ttctgaagg  
 6061 aactgttggg tgggtcaca ttcaggctat atattccca gggctcagcc agtggatcaa  
 6121 tcaatcatc cgtgacttct tctgtgttgg tgttaacctata tctataattt  
 6181 cgtttattaa aatttaatat atttcgacga tgaatttctc aaggatattt ttcttctgt  
 6241 tctgttgggt tctgtgttgg tcaacagttt cggctcggcc agagcgaaa ggtaccag  
 6301 tgcagctga ggaagtcggg ggaagcttgg taaagccggg ggggtccctt agagtctct  
 6361 tgcagctc tggatttact ttgagaaacg cttggatgag ctgggtccgc caggctccag  
 6421 ggaagggggt ggaattgggt ggcgtatta aagcaaaat tgaattggg acaacagact  
 6481 atgtgtgacc cgtgaaaggc agattcaca tctcaagaga tgaattcaaa aacagttat  
 6541 atctgcaaat gaatagctg aagcccgagg acacgctgt atattactgt accacgggga  
 6601 ttatgataac atttggggga gttatccctc cccgaattg gggccaggga accctgggtca  
 6661 cgtctctc agcctccacc aagggtccat cgtctctcc cctggcacc tctccaaga  
 6721 gacacctgg gggcacagcg gccctgggt gccctgggtcaa ggaactactt cccgaaccgg  
 6781 tgcagctgtc tgggaactca ggcggccctga ccagcgjcg gcacacctt cggctgttcc  
 6841 tacagcttc aggaactac ttcttagca acgtgaatc aagcccgag caacccaa gtggacaaga  
 6901 gacccagac ctacatctgc aacgtgaatc aagcccgag caacccaa gtggacaaga  
 6961 aggtgagcc caaatctgt gacaaactc acacatgcc accgtgcca gcactgaac  
 7021 tctgggggg accgtcagtc ttctcttcc cccaaaacc caaggacacc ctcatgatct  
 7081 cccggacccc tgaagtcaca tgcgtgttgg tgaagctgag ccagaagac cctgaggtca  
 7141 agttcaactg gtacgtggac ggcgtggagg tgcataatgc caagacaaag cgcggggagg  
 7201 agcagtaaca cagcagctac cgtgtgttca ggtcctcac cgtcctgac caggactggc  
 7261 tgaatggcaa ggaatgaaag tgaaggtct caacaaagc cctccagcc ccatcgaga  
 7321 aacacatctc caagccaaa gggcagcccc gagaaccaca ggtgtacacc ctgccccat  
 7381 cccgggatga gctgaccaa accaggtca gctgacctg cctgggtcaa ggtctctatc  
 7441 ccagcgacat cgcgtggag tggagagga atgggcagcc ggagaaacac tacaagacca  
 7501 cgtctccctg gctggactcc gacggtctct tcttctcta cagcaagctc accgtggaca  
 7561 agacaggtg gcagcagggg aacgtcttct catgtcctgt ctcgggtaa agcgcagag ccgaaaaagc  
 7621 accactaac gcagagagc ctctccctgt ctcgggtaa agcgcagag ccgaaaaagc  
 7681 ttctctatga gctgacacag ccacctcgg tgtcagtgct ccagggaca accggccagg  
 7741 tcaactgctc tggagatgca ttgacagaa aatatgttta ttgggtaccag cagaagtcag  
 7801 gcaagggccc tgtgtgtgtc atctatgag accgaaacg accctccggg atccctgaga  
 7861 gattctctgg tccagctca gggacaatgg ccacttgac tatcagtggg gccagggtgg  
 7921 aagatgaagg tgactactac accgtctag gtcagccaa ggtgcccc ttggtcactc  
 7981 tcagcggagg gacaaagctg accgtctag gtcagccaa ggtgcccc ttggtcactc  
 8041 tgttccacc ctctctgag gacttcaag ccaacaggc cacactgtg tgtctcataa  
 8101 gtgactccta cccgggagcc gtgacagtgg cctggaggc agatagagc ccgtcaagg  
 8161 cgggagtgga gaccaccaca ccctccaaac aagcaaca caagtaacg gccagcagct  
 8221 acctgacct gacgttgag cagtgaagt ccccaaaag ctacagctgc caggtcacgc  
 8281 atgaagggag caccgtggag aagacagtgg cccctgcaga atgttcaccg cggaggagg

5

10

15

20

25

30

35

8341 gaaggggccct ttttgaaggg ggaggaact tcgggccatg actcctctcg tgccccccgc  
 8401 acggaacact gatgtgcaga gggccctctg ccattgctgc ttccctctgccc cttccctgctc  
 8461 actctgaatg tggcttcttt gctactgcca agcagtgctt tccccagca ccccttgctg  
 8521 ggggttctctg agatttttca agagtgcgtt agcacattcc tccccagca ccccttgctg  
 8581 caggccagtg ccaggcaccg acttggctac tgcctgcccc gagagaaatc cagttcaata  
 8641 ttttccaaa ccaaatggat tacaatgccc ctgactgccc attaacaggt gtttgcattt  
 8701 atctgtgctt tcgcttcacc cacattatcc cattgcctcc cctcgagggg gggcccggtg  
 8761 cccaattcgc cctatagtga gtcgtattac gcgcgtcac tggcgtcgtt tttaaacgtt  
 8821 cgtgactggg aaacctctgg cgttaccgca cttaatcgcc ttgcagcaca tccccctttc  
 8881 gccagctggc gtaactgcga agaggcccgcc accgatcgcc cttccaaca gttggcgagc  
 8941 ctgaatggcg aatggaatt gtaagcgtta atatttgggt aaaattcgcc ttaattttt  
 9001 gttaatcag ctcatctttt aaccaatagg ccgaatcggt caaatccct tataaatcaa  
 9061 aagaatagac cgagataggg ttgagtgttg ttccagtttg gaacaagagt ccactattaa  
 9121 agaactgga cttcaactgc aaaggcgaa aaaccg+cta tcaggggcgat ggcactactac  
 9181 tccgggatca tatgacaaga tgtgtatcca ccttaactta atgattttta ccaaatcat  
 9241 taggggattc atcagtgtc aggttcacg agaattaaca ttccgtcagg aaagcttatg  
 9301 atgagtgtg gcttaaaaac ttactcaatg gctgggttatg catatgcgaa tacatgcgaa  
 9361 aaacttaaaa gagcttgccg ataaaaagg ccaatttatt gctatttacc ggggttttt  
 9421 attgagcttg aagataaaat aaatagata ggttttattt gaagctaaat cttctttatc  
 9481 gtaaaaaatg cctcttggg ttatcaagag ggtcattata ttctgcgaa taacatcat  
 9541 tggtagcga ataaactagc actgtctccc tgtttactcc cctgagcttg aggggttaac  
 9601 atgaaggtca tcatagcag gataataata cagttaaacg ctaaaccaat aatccaaatc  
 9661 cagccatccc aaattggtag tgaatgatta taaataacag caaacagtaa tggggccaata  
 9721 acaccgggtg catctggtag gctcaccat aatccctgta agcaccttg ctgatgctc  
 9781 ttgttttga tagacatcac tccctgtaat gcaggtaaaag cgatcccacc accag+caat  
 9841 aaaattaaa cagggaatac taocaaacct gagcagttact gcgctttttt gcccattta gtcgctattc  
 9901 ctactatctg caataaatcc gagcagttact gcgctttttt gcccattta gtcgctattc  
 9961 ttcctgccc aaaggcttgg aatactgagt gtaaaagacc aagaccgta atgaaaagcc  
 10021 aaccatcatg ctattcatca tcaagatttc tgaatagca ccacacgtg ctggattggc  
 10081 tatcaatgcy ctgaaataat aatcaacaaa tggcatcggt aaataagtga tgtatacga  
 10141 tcagcttttg tcccccttag tgaagggttaa ttgocgctt ggcgtaatca tggctatagc  
 10201 tgtttcctgt ggaatattgt tatccgctca caattccaca caacatagca gccggaagca  
 10261 taaagtgtaa agcctgggtt gcctaattgag tgaactaact cacattaatt gcgttgctt  
 10321 cactgcccgc ttccagctcg ggaacacctg cgtgccagct gcattaatga atcgcccaac  
 10381 gcgcggggag aggcgggttg cgtattggcg gctcttcgc ttctcgtc actgactcgc  
 10441 tgcgtcgggt cgttcgctg cgcgcagcgg tatcagctca ctcaaggcg gtaatacgg  
 10501 tatccacaga atcagggggt aacgcaggaa agaactgtg agcaaaaggc cagcaaaagg  
 10561 ccagggaaccg taaaaggcc ggttgctgg cgtttttcca taggtccgc cccctgacg  
 10621 agcatcaca aaatcgacgc tcaagtacga ggtggcgaaa ccgacagga ctataaagat

5

10

15

20

25

30

35

10681 accaggcggtt tccccctgga agctccctcg tgcgctctcc tgttcgacc ctgcccgttta  
 10741 cgggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat agctcacgct  
 10801 fraggatatt cagttcgggtg tagtgcgttc gctcaagct gggctgtgtg cagaaacccc  
 10861 cgtttcagcc cgaccgctgc gccctatccg gtaactatcg tcttgagtc aaccggtaa  
 10921 gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga gcgagggtatg  
 10981 tagggcgtgc tacagagttc ttgaagtggg ggcctaacta cggctacact agaaggacag  
 11041 tatttgggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt  
 11101 gatccggcaa acaaacacc accggttagcg gtggtttttt tgtttgcaag cagcagatta  
 11161 cgcgcagaaa aaagcagct caagaagatc cttgatctt ttctacgggg tctgacgctc  
 11221 agtggaaaca aaactcagct taagggattt tggctcatgag attatcaaaa aggatcttca  
 11281 cctagatcct tttaaatcaa aatgaagtt ttaaatcaat cttaagata tatgagtaaa  
 11341 cttgggtctga cagttaccaa tgccttaata gtgagggacc tatctcagcg atctgtctat  
 11401 ttcggttcac catagttgct tgactcccg tctgttagat aactacgata cgggagggct  
 11461 taccatctgg cccagtgct geatgtatc cgcgagccc acgtccacc gctccagatt  
 11521 tatcagcaat aaaccagcca gcgggaaggg ccgagcgag agtggtcct gcaactttat  
 11581 ccgcctccat ccagttctat aattgttgcc ggaagctag agtaagtagt tcgccagtta  
 11641 atagtttgcg caacgttgtt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg  
 11701 gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga tccccatgt  
 11761 tgtgaaaaa agcgggttagc tctctggtc ctccagatcgt tgcagaagt aagttggccg  
 11821 cagtggtatc actcatggtt atggcagcac tgcataaatt tcttactgtc atgccatccg  
 11881 taagatgctt ttctgtgact ggtgagtagt caaccaagtc attctgagaa tagtgtatgc  
 11941 ggcgaccgag ttgctcttgc cggcgctcaa tacgggataa tacgcgcca catagcagaa  
 12001 ctttaaaagt gctcatcatt ggaatacgtt cttcggggcg aaactctca agtatcttac  
 12061 cgctgttgag atccagttcg atgtaaccca ctctgacacc caactgatct tcagcatctt  
 12121 ttactttcac cagcgtttct gggtaggcaa aaacagggaag gcaaaatgcc gcaaaaaagg  
 12181 gaataaggcg gacacggaaa tgttgaatc tcatactctt cctttttcaa tattattgaa  
 12241 gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata  
 12301 aacaaatagg ggttcgcgcg acatttcccc gaaaagtgc ac

5  
 10  
 15  
 20  
 25  
 30

SEQ ID NO:104 (pTnMOD (CMV-prepro-HCPro-CPA))  
 1 ctgacgcgc ctgtagcggc gcattaaagc cggcgggtgt ggtgggttac cgacgcgtga  
 61 ccgtacact tgcacgcgc ctacgcccc ctctttcgc ttcttccct tctttctcg  
 121 ccaggttcgc cggcatcaga ttggctattg gccattgat aggttgatc catatcataa  
 181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac  
 241 tagttattaa tagtaataca ttacgggggtc attagtcat agcccatata tggagttccg  
 301 cgttacataa cttacgggtaa atggcccgcc tggctgaccg cccaacgacc ccgcccatt  
 361 gagtcaata atgacgtatg ttcccatagt aacgcaata gggactttcc attgacgtca  
 421 atgggtggag tattacggt aaactgccc cttggcagta catcaagtgt atcatatgcc  
 481 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta

35



541 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgtattac  
 601 catgggtgat cggttttggc agtacatcaa tgggcgttga tagcgttttg actcagggg  
 661 atttccaagt ctccacocca ttgagctcaa tgggagtttg ttttggcacc aaatcaacg  
 721 ggactttcca aaatgtcgta acaactcgc ccattgacg caaatggcg gtagcgtgt  
 781 acggtgggag gtctatataa gcagagctcg tttagtgac cgtcagatcg cctggagacg  
 841 ccattccacg tgttttgacc tccatagaag acacgggac cgtccagcc tcgcggccg  
 901 ggaacggtgc attggaacgc ggattcccg tgccaagagt gacgtaahta ccgctatatg  
 961 actctatagg cacaccctt tggctcttat gcatgctata ctgtttttg cttggggcct  
 1021 atacacccc gcttctttat gctatagggt atggatatag ttagcctata ggtgtgggtt  
 1081 attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac  
 1141 atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac  
 1201 tgacacggac tctgtatttt tacaggatgg ggtccatttt attatttaca aattcacata  
 1261 tacaacaacg ccgtccccc tgcccgcatg ttttattaaa catagcgtgg gatctccag  
 1321 cgaatctcgg gtacgtgttc cggacatggg ctcttc\*ccg gtacggcggt agcttccaca  
 1381 tccagacctt ggtcccatgc ctccagcggc tcatggtcgc tggcagctc cttgtctcta  
 1441 acagtggagg ccagacttag gcacagcaca atgccacca ccaccagtgt gcgcacaaag  
 1501 gccgtggcgg tagggtatgt gctgaaaat gagcgtggag attgggtcgc cagcgtgac  
 1561 gagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtattc  
 1621 tgataagagt cagaggtaac tccggttgcg gtgctgttaa cgggtggagg cagtgtagtc  
 1681 tgagcagtag cagttgtcgc cgcgcgcgc accagacata atagctgaca gactaacaga  
 1741 ctgttccctt ccatgggtct tttctgcagt caccgtcggg ccatgtgtga acttgatat  
 1801 ttacatgatt ctctttacca attctgccc gaattacact taaacgact caacagctta  
 1861 acgttggctt gccacgcat acttgactgt aaactctca ctctaccga acttggcgt  
 1921 aacctgccaa ccaagcggag acaaaaacat acatcaaac gaatcgaccg attgtt\*agt  
 1981 aatcgtcacc tccacaaaga gcgactcgt gtataccgtt ggcatgctag ctttalctgt  
 2041 tccgggaata cgtatgccat tgtacttgtt gactggtctg atattcgtga gcaaaaacga  
 2101 ctataggtat tgcgagcttc agtcgcacta caggttcgtt ctgttactct ttatgagaaa  
 2161 gcgttccgc tttcagagca atgttcaaag aaagctcag accaatttct agccgacctt  
 2221 gcgagcattc taccgagtaa caccacacgc ctcaattgca gtatgctgg ctttaaagt  
 2281 ccattggtata aatccgttga gaagctgggt tgggtactgt taagtctagt aagaggaaaa  
 2341 gtacaatatg cagacctagg agcggaaaaac tggaaacctc tcagcaactt acatgatag  
 2401 tcatctagtc actcaaaagc tttaggtctat agaggtctga ctaaaagcaa tccaatctca  
 2461 tggcaaatc tattgtataa atctcgtctt aaagggcga aaatcagcg ctgcacacg  
 2521 actcattgtc accaccgtc acctaaaac tactcagct cggcaaaaga gccatgggtt  
 2581 tagcaacta acttacctgt tgaattcga acccttgaac gtccctgcta cgtactatcg  
 2641 aaggaatgc agattgaaga aacctccga gacttgaata tcatgtctat cggactaggc  
 2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgtctat aatcgccctg  
 2761 atgtctcaac taacatgttg gcttgcgggc gttcatgctc agaacaaggt ttgggacaa  
 2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg ctttaggcagt

5

10

15

20

25

30

35

2891 gaagtttttc ggcattcttg ctacacaata acaagggaag acttactcgt ggctgcaacc  
 2941 ctactagctc aaattttatt cacacatggt tacgcttttg ggaattatg ataataatcc  
 3001 agatcacttc tggtaataa aagatcacag ctctagagat ctgtgtgttg gttttttgtg  
 3061 gatctgctgt gctttctagt tgcagccat ctgtgtgttg cccctcccc gtgccttctt  
 3121 tgaccttga aggtgccact cccactgtcc ttctctaata aaatgaggaa attgcatcgc  
 3181 attgtctgag taggtgtcat tctattcttg ggggtggggg ggggcagcac agcaaggggg  
 3241 aggattggga agacaatagc aggcattgct gggatgcggt gggctctatg ggtacctctc  
 3301 tctctctctc tctctctctc tctctctctc tctctctctc tctctctctc tctctctctc  
 3361 tctctctctc tctctctctc tctctctctc tctctctctc tctctctctc tctctctctc  
 3421 tgccttttat catcacttta aaataaaaa acaattactc agtgctgtt ataagcagca  
 3481 attaatatg attgatgcct acatcacac aaaaactgat ttaacaaatg gtgtgtctgc  
 3541 cttagaaagt atatttgaac attatcttga ttatattatt gataataata aaaacttat  
 3601 ccctatccaa gaagtgtgc ctatcatttg ttggaatgaa ctgaaaaaaa attagccttg  
 3661 aacatactac tggtaaggta aacgccattg tcagcaaat gatcaagag aaccaactta  
 3721 aagctttctc gacggaatgt taattctcgt tgaccttgag cactgatgaa tccctaattg  
 3781 attttggtaa aaatcattaa gttaaagtgg atacacatct tgcataatga tccgggtaat  
 3841 gtgagttagc tcaactatta ggcacccacg gctttacact ttatgcttcc ggtcgtatg  
 3901 ttgttggtaa ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac  
 3961 gccagcgcg caattaaccc tcaataaagg gaacaaaagc tggagctcca cggcggtggc  
 4021 ggcgctcta gaactagtgg atccccgggg ctgcaggaat togatatacaa gcttatcgat  
 4081 accgttgacc tgcagcatca gattggctat tggccattgc atcgttgtta tccatatcat  
 4141 aatatgtaca ttatatttgg ctcatgtcca acattacgc catgttgaca ttgattattg  
 4201 actagtatt aatagtaatc aattacgggg tcatgttct atagccata tatggagttc  
 4261 cgcgttacat aacttacggt aaatggccc cctggctgac ggcacaaaga ccccccca  
 4321 ttgacgtcaa taatgacgta tgttccata gtaacgcaa tagggacttt ccatgacgt  
 4381 caatgggttg agtatttac gtaactgcc cacttgccag tacatcaagt gtatcatatg  
 4441 tcaagtacgc cccctattga cgtcaatgac ggtaaatggc ccgctggca ttatgcccag  
 4501 tacatgacct tatgggactt tctacttgg cagtacatct acgtattagt catcgtatt  
 4561 accatgggtga tgcggttttg gcagtacatc aatgggctg gataggggtt tgactcacgg  
 4621 gcatttccaa gcttccacc catgacgtc aatgggaggt tgttttgga ccaaatcaa  
 4681 cgggactttc caaatgctg taacaactcc gcccatgga cgttaagtga cgttaggggt  
 4741 gtacgggtgg aggtctatat aagcagagct cgttaagtga accgtcagat cgcctggaga  
 4801 cccatccac gctgttttga cctcataga agacacgggg accgatccag cctccgggc  
 4861 cgggaacgggt gcaattggaac gggattccc cgtgcaaga gtgacgtaag taccgctat  
 4921 agactctata ggcacacccc ttgggtctt atgatgcta tactgttttt ggctggggg  
 4981 ctatacaccc cggcttctt atgctatagg tgatgggata gcttagccta taggtgtggg  
 5041 ttattgacca ttattgacca ctccctatt ggtgacgata ctttccatta ctaatacata  
 5101 acatggctct ttgccacaac tatctctatt ggttatatgc caatactctg tccctcagag  
 5161 actgacacgg actctgtatt ttacaggat ggggtcccat ttattattta caaattcaca

5

10

15

20

25

30

35

5221 tatacaacaa cgccgtcccc cgtgccccga gtttttatta aacatagcgt gggatctcca  
 5281 cgcgaatctc gggtagctgt tccggacatg ggtcttcttc cggtagcggc ggagcttcca  
 5341 catccagacc ctgggtcccat gctccagcg gtcctatggtc gctcggcagc tccttgctcc  
 5401 taacagtggg gggcagactt aggcacagca caatgcccc caccaccagt gtgcgcgaca  
 5461 aggcgtggc ggtagggtat ggtctgtaaa atgagcgtgg agattgggct cgcacggctg  
 5521 acgcagatgg aagacttaag gcagcggcag aagaagatgc aggcagctga gttgtgtat  
 5581 tctgataaga gtcagaggtta actcccgtt cccgcgcgcg caaccagaca taatagctga cagactaaca  
 5641 tctggcagct actcgttctt tccatgggt cttttctgca gtcaccgtcg gatcaatcat tcctctcgtg  
 5701 gactgttctt tccatgggt tacctatata tctaaattta atatttcgtt tattaaaaatt  
 5761 actttctcgt gtcagatgaa tttctcaagg atattttctc tegtgttcgc tttggttctg  
 5821 taatatattt cgaagatgaa tttctcaagg atattttctc tegtgttcgc tttggttctg  
 5881 gctttgtcaa cagtttctgc tgcgcccag cgaagaggtta cccaggtgca gctgcaggag  
 5941 tcgggggggag gcttggtaaa gccggggggg tcccttagag tctcctgtgc agcctctgga  
 6001 ttcactttca gaacgcctg gatgagctgg gtccgcagg ctcacaggaa ggggctggag  
 6061 tgggtcggc gatttaaaag caaaattgat ggtgggacaa cagactatgc tgcaccctg  
 6121 aaggcagat tcaccatctc aagagatgat tcaaaaaaca cgttatatct gcaaatgaat  
 6181 agcctgaaag ccgaggacac agcgttatat tactgtacca cggggatttat gataacattt  
 6241 gggggagtta tccctcccc gaattggggc cagggaaccc tggtcacgt ctctcagcc  
 6301 tccaccaagg gccatcgtt cttccccctg gcacctctct ccaagagcac ctctggggg  
 6361 acagcggccc tgggtgctt ggtcaaggac tacttcccc aaccggtgac ggtgctgtgg  
 6421 aactcaggcg ccttgaccag cggcgtgcac accttcccc ctgtcctaca gctcctcaga  
 6481 cttacttcc tttagcaact ggtgacctg cctccagca gcttgggcac ccagacctac  
 6541 atctgcaacg tgaatcaca gccagcaac accaagggtg acaagaaagt tgagcccaaa  
 6601 tcttgtgaca aaactcacac atgcccaccg tgcacagcac ctgaactctt ggggggaccg  
 6661 ttagtcttcc tcttcccccc aaaccccaag gacacctca tgatctccc gaccttgag  
 6721 gtcacatgcy tgggtggtgga cgtgagccac gaagacctg aggtcaagtt caactggtac  
 6781 gtggacggcg tggagggtgca taatgccaaag acaagccgc gggaggagca gtacaacagc  
 6841 acgtaccgtg tggtcagctt cctcaccgtc ctgacccagg actggtgaa tggcaaggag  
 6901 tacaagtga aggtctccaa caaagccctc ccagccccc tcgagaaaaac catctccaaa  
 6961 gccaaaggcg agccccgaga accacagggtg tacacctgc cccatcccc ggatgagctg  
 7021 accaagaacc aggtcagctt gacctgctg gtcacaggct tctatcccc cgacatcgcc  
 7081 gtggagtggg agagcaatgg gcagccggag acaactaca agaccacgoc tccgtgctg  
 7141 gactcagacg gctccttctt cctctacagc aggtcaccg tggacaagag caggtggcag  
 7201 caggggagcg tcttctcatg tcccgtagtg catgaggctc tgcacaacca ctacacgag  
 7261 aagagctctt cctgtctcc ggttaagcg ccagagccga agcttctcta tgagctgaca  
 7321 cagccacctt cgggtgtagt gtccccagga caaacggcca ggtacacctg ctctggagat  
 7381 gcatgcccag aaaaatatgt ttattggtac caggccaggc cctgtgggtg  
 7441 gtacttatg aggcagacaa acgacctcc gggatcccc agagattctc tggctccagc  
 7501 tcagggacaa tggccacctt gactatcagt gggggccagg tggagatga aggtgactac

5

10

15

20

25

30

35

7561 tactgttact caactgacag cagtgggttat cataggggagg tgttcagcgg agggaccaag  
 7621 ctgacggtcc taggtcagcc caaggctgcc cctcgggtca ctctgttccc accctcctct  
 7681 gaggagcttc aagccaacaa ggcacactg gtgtgtctca taagtgactc ctaccgggga  
 7741 gccgtgacag tggcctggaa ggcagatagc agccccgtca aggcgggagt ggaaccacc  
 7801 acacctcca acaaaagcaa caacaagtac gcggccagca gctacctgag cctgacgctt  
 7861 gagcagtga agtccacaa aagctacagc tggcagggtca cgcataaagg gacaccgtg  
 7921 gagaagacag tggccctgc agaatttca cgcggagggg agggaaaggg cctttttgaa  
 7981 gggggaggaa acttcgcgc atgactcctc tegtgcgcc cgcacgggac cctttttgaa  
 8041 agagggccct ctgccattgc tgcctcctct gcccttctc gtcactctga atgtggcttc  
 8101 ttgtctactg ccacagcaag aaataaaatc tcaacatcta aatgggtttc ctgagatttt  
 8161 tcaagagtog ttaagcacat tccctcccca gcacctctg ctgcaggcca gtgccaggca  
 8221 ccaacttggc tactgtgcc catgagagaa atccagttca atattttcca aagcaaaatg  
 8281 gattacatat gccctagatc ctgattaaac ggtgttttgt attatctgtg ctttgccttc  
 8341 accacatta tccattgcc tccctcgac tcgagggggg gcccggtacc caattgcgcc  
 8401 tatagttagt cgtattacgc gogtccactg gccgtcgttt tacaagtcg tgaactggaa  
 8461 aaccttgccg ttaccacact taatgcctt gcagcacatc ccccttctgc cagctggcgt  
 8521 aatagcgaag agggccgcac cgatcgccct tcccaacagt tgcgagcct gaatggcgaa  
 8581 tggaaattgt aagcgttaatt attttgttaa aattcgcgtt aaattttgt taaatcagct  
 8641 ctttttttaa ccaataggcc gaattcggca aatccctta taaatcaaaa gaatagaccg  
 8701 agataggggt gagtgttgtt ccagtttggg acaagagtcc actattaaag aacgtggact  
 8761 ccaacgtcaa agggcgaaaa accgtctatc agggcgatgg cccactactc cgggattcata  
 8821 tgacaagatg tgtatccacc ttaacttaat gatitttacc aaatcatta ggggattcat  
 8881 cagtgtcag ggtcaacgag aattaacatt ccgtcaggaa agcttatgat gatgatgtgc  
 8941 ttaaaaactt actcaatggc tggttatgca tctcgcaata catcgaaaa acctaaaaa  
 9001 gcttgcgat aaaaaaggcc aatttattgc tatttacgc ggtttttat tgacctgaa  
 9061 agataaataa aatagatagg ttttatttga agctaaatct tctttatcgt aaaaaatgcc  
 9121 ctcttgggtt atcaagaggg tcattatatt tgcgcggaata acatcatttg gtgacgaaat  
 9181 aactaagcac ttgtctcctg tttactcccc tgagcttgag ggtttaacat gaaggtcatc  
 9241 gatagcagga taataatata gtaaaacgct aaaccaataa tccaaatcca gccatcccaa  
 9301 attggttagt aatgattata aataacagca aacagtaatg ggccaataac accggttgca  
 9361 ttggttaagg tcaccaataa tccctgtaaa gcaccttgct gatgactctt tgtttggata  
 9421 gacatcactc cctgtaatgc aggtaaagcg atcccaaac gctaaaaagg caaatgact actatctgca  
 9481 ggaataacta accaaccttc agataaac gctaaaaagg gctattatgt ggctattctt cctgcacaa  
 9541 ataaatcoga gcagtactgc cgttttttgc cccatttagt ggtatttctt cctgcacaa  
 9601 aggtttgaa tactgagtgt aagagacca gaccgttaac gaaagccaa ccatcatgct  
 9661 attcatcatc acgattttct taatgacac acacgtgctt ggattggcta tcaatggct  
 9721 gaaataataa tcaacaaatg gcacgttaa ataatgtatg tatacogact agcttttgtt  
 9781 ccttttagtg aggtttaatt gcggccttgg cgtaatcatg gtcatagctg ttctctgtgt  
 9841 gaaattgtta tccgctcaca attccacaca acatacagc cggaagcata aagtgtaaaag

5

10

15

20

25

30

35

9901 cctggggtgc ctaatgagtg agtaactca cattaattgc gttgcgctca ctgcccggctt  
 9961 tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaaagc gcggggagag  
 10021 gcgggttgcg tattgggcgc tcttcgcgtt cctgcgtcac tgactcgctg cgtcgggtcg  
 10081 ttccgggtgcg gcgagcggta tcagctcact caaaggcggg aatacggtta tccacagaat  
 10141 caggggataa cgcaggaaa aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta  
 10201 aaaaggccgc gtgtcggcg tttttccata ggtccgcgc cctgacgag catcacaaa  
 10261 atcagcgtc aagtcagagg tggcgaaacc cagacgact ataaagatac caggcgtttc  
 10321 cccctgggaag ctccctcgtg cgctcctcgt ttcacacctt gcgcttacc ggatacctgt  
 10381 cgcctttct ccttcggga agcgtggcg tttctcatag ctacgctgt aggtatctca  
 10441 gttcgggtga ggtcgtcgc tccaagctgg gctgtgtgca cgaaccccc gttcagccg  
 10501 accgctgcgc cttatcgggt aactatcgtc ttgagtcaca ccgggtaaaga caagacttat  
 10561 cgccactggc agcagccact ggtaacagga ttacgacagc gaggtatgta ggcggtgcta  
 10621 cagagttctt gaagtgggtg cctaactacg gctacactag agggacagta ttggtatct  
 10681 gcgtctcgt gaagcagtt accttcggaa aaagagtgg tagctcttga tccggcaaac  
 10741 aaaccaccgc tggtagcggg ggtttttttg ttgcaagca gaagattacg cgcagaaaaa  
 10801 aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaaacgaaa  
 10861 actcactgta agggattttg gtcattgagat tatcaaaaag gatcttcacc tagatccttt  
 10921 taattataaa atgaagtttt aaatcaatct aaagtataata tgagtataact tggcttgaca  
 10981 gttaccaatg cttaatcagt gagcaccta tctcagcgat ctgtctattt cgttcattca  
 11041 tagttgcctg actcccgctc gtgtagataa ctacgatacg ggagggtcta ccatctggcc  
 11101 ccagtgctgc atgataccg cgagaccac gctcacggc tccagattta tcagcaataa  
 11161 accagccagc cggaaaggcc gagcgcagaa gtggctcgc aactttatcc gcctccatcc  
 11221 agtctattaa ttgttgccgg gaagctagag taagtgttc gccagttaat agtttgcgca  
 11281 acgtgtgtgc cattgctaca ggcacgtggtg tgtcacgctc gtcgtttggg atgggttcat  
 11341 tcagctcggg ttcccaacga tcaaggcggag ttacatgac cccatgttg tgcaaaaag  
 11401 cggttagctc cttcgggtcct ccgacgtgtg tcagaagtaa gttggccgca gtgttatcac  
 11461 tcatggttat ggcagcactg cataatttc ttaactgcat gccatccgta agatgctttt  
 11521 ctgtgactgg tgagtactca accaagtcat tctgagaata gtgtatcgg cgaccgagtt  
 11581 gctcttgccc ggcgtcaata cgggataata ccgcccaca tagcagaact ttaaaagtgc  
 11641 tcatcatbgy aaacgtttct tcggggcgaa aactcgaag gatcttaccg ctgttgagat  
 11701 ccagttcagat gtaaccact cgtgcacca actgactctt agcatctttt actttacca  
 11761 ggtttctcgg gtgagcaaaa acaggaaggc aaatggccg aaaaaaggga ataaaggcga  
 11821 cagggaaatg ttgaatactc atactcttc ttttcaata ttattgaagc atttatcagg  
 11881 gttattgtct catgagcggg tacatatgtg aatgrattta gaaaaataaa caaatagggg  
 11941 ttccgcgcac atttcccga aaagtgcac

## CLAIMS

We Claim:

1. An isolated polynucleotide comprising at least one gene of interest and one or more pro nucleotide sequences, wherein the at least one gene of interest is operably-linked to a pro nucleotide sequence and each of the at least one gene of interest may be the same or different when more than one one gene of interest is present, wherein the at least one gene of interest encodes a protein useful in a vaccine.
2. The polynucleotide of claim 1, wherein a most 5' pro nucleotide sequence of the one or more pro nucleotide sequences is a part of a prepro nucleotide sequence.
3. The polynucleotide of claim 1, wherein two genes of interest and two pro nucleotide sequences are arranged in the following order: a prepro nucleotide sequence, a first gene of interest, a pro nucleotide sequence, and a second gene of interest.
4. The polynucleotide of claim 3, wherein the prepro nucleotide sequence is a cecropin prepro nucleotide sequence and the pro nucleotide sequence is a cecropin pro sequence.
5. The polynucleotide of claim 3, wherein the prepro nucleotide sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4 and the pro nucleotide sequence comprises a sequence shown in SEQ ID NO:1 or SEQ ID NO:2.
6. The polynucleotide of claim 1, wherein a first gene of interest encodes for an antibody heavy chain and a second gene of interest encodes for an antibody light chain.
7. A method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide comprising two or

more genes of interest, wherein each gene of interest encodes a part of the multimeric protein, each gene of interest is operably-linked to a pro nucleotide sequence, and each of the two or more genes of interest may be the same or different, wherein each of the two or more genes of interest encodes a protein useful in a vaccine.

5

8. The method of claim 7, wherein a most 5' pro nucleotide sequence of the two or more pro sequences is a part of a prepro nucleotide sequence.

10

9. The method of claim 8, wherein the polynucleotide comprises two genes of interest and two pro nucleotide sequences arranged in the following order: a prepro nucleotide sequence, a first gene of interest, a pro nucleotide sequence, and a second gene of interest.

15

10. The method of claim 7, wherein a first gene of interest encodes for an antibody heavy chain and a second gene of interest encodes for an antibody light chain.

20

11. A method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide comprising a cecropin prepro nucleotide sequence operably-linked to two or more genes of interest, each gene of interest encoding a part of the multimeric protein, wherein the multimeric protein is useful in a vaccine.

25

12. The method of claim 11, wherein a first gene of interest is an antibody heavy chain and a second gene of interest is an antibody light chain.

30

13. A method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide comprising two or more genes of interest, wherein each gene of interest encodes a part of the multimeric protein and wherein each gene of interest is operably linked to a gene encoding for a cleavage site, wherein the multimeric protein is useful in a vaccine.

14. The method of claim 13, wherein a transposon-based vector comprises the polynucleotide and further comprises a transposase gene operably linked to a first promoter and wherein;

5           a) the first promoter comprises a modified Kozak sequence comprising ACCATG;

          b) the two or more genes of interest are each operably-linked to one or more additional promoters; and,

10           c) the two or more genes of interest and their operably-linked promoters are flanked by transposase insertion sequences recognized by a transposase encoded by the transposase gene.

15           15. The method of claim 13, wherein a transposon-based vector comprises the polynucleotide and further comprises a transposase gene operably linked to a first promoter and an avian optimized polyA sequence, and wherein;

          a) the two or more genes of interest are each operably-linked to one or more additional promoters; and,

20           b) the two or more genes of interest and their operably-linked promoters are flanked by transposase insertion sequences recognized by a transposase encoded by the transposase gene.

16. An animal comprising the isolated polynucleotide of Claim 1.

25           17. The animal of Claim 16, wherein the animal is a bird.

18. An egg produced by the animal of Claim 16.

30           19. The egg of Claim 18, wherein the egg comprises a multimeric protein encoded by the isolated polynucleotide.

20. The animal of Claim 16, wherein the animal is a mammal.

21. Milk produced by the mammal of Claim 20.



22. The milk of Claim 21, wherein the milk comprises a multimeric protein encoded by the isolated polynucleotide.
- 5 23. A method of producing a multimeric protein comprising:
- a) administering to an egg-laying animal a composition comprising the polynucleotide of Claim 1; and,
  - b) permitting the one or more genes of interest to be expressed into the multimeric protein, wherein the multimeric protein is useful in a vaccine.
- 10
24. The method of Claim 23, further comprising
- a) collecting an egg from the egg-laying animal;
  - b) harvesting egg white comprising the multimeric protein; and,
  - c) purifying the multimeric protein.
- 15
25. The method of Claim 23, wherein the egg-laying animal is a bird.
26. A method of producing a multimeric protein comprising:
- a) administering to an intramammary duct system of a mammal a composition comprising the polynucleotide of Claim 1, and,
  - b) permitting the one or more genes of interest to be expressed into the multimeric protein, wherein the multimeric protein is useful in a vaccine.
- 20
27. The method of Claim 26, further comprising
- a) collecting milk from the mammal, wherein the milk comprises the multimeric protein; and
  - b) purifying the multimeric protein.
- 25
28. A method of vaccinating an individual against a disease comprising administering to the individual a composition comprising a protein, a multimeric protein or an antibody or a combination thereof, wherein the
- 30

composition is produced in the individual or in another individual made transgenic using transposon-based vectors.

5           29.     Use of a composition comprising a protein, a multimeric protein or an antibody or a combination thereof, wherein the composition is produced in an individual made transgenic using transposon-based vectors, in the preparation of a medicament for use in a vaccine.

10           30.     The polynucleotide of any of the preceding claims, wherein the prepro nucleotide sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4.

15           31.     The polynucleotide of any of the preceding claims, wherein the two or more pro nucleotide sequences each comprise a sequence shown in SEQ ID NO:1 or SEQ ID NO:2.

            32.     The polynucleotide of any of the preceding claims, wherein the prepro nucleotide sequence is a cecropin prepro nucleotide sequence.

20           33.     The polynucleotide of any of the preceding claims, wherein the pro nucleotide sequence is a cecropin pro nucleotide sequence.

FIGURE 1

Prom	prepro	Gene of Interest	pro	Gene of interest	polyA
------	--------	------------------	-----	------------------	-------

**THIS PAGE BLANK (USPTO)**

FIGURE 2

IS	Oval Prom	prepro	Heavy chain	pro	Light chain	polyA	IS
----	-----------	--------	-------------	-----	-------------	-------	----

**THIS PAGE BLANK (USPTO)**

FIGURE 3

Prom	Cecropin prepro	Gene of Interest	CS	Gene of Interest	polyA
------	-----------------	------------------	----	------------------	-------

**THIS PAGE BLANK (USPTO)**

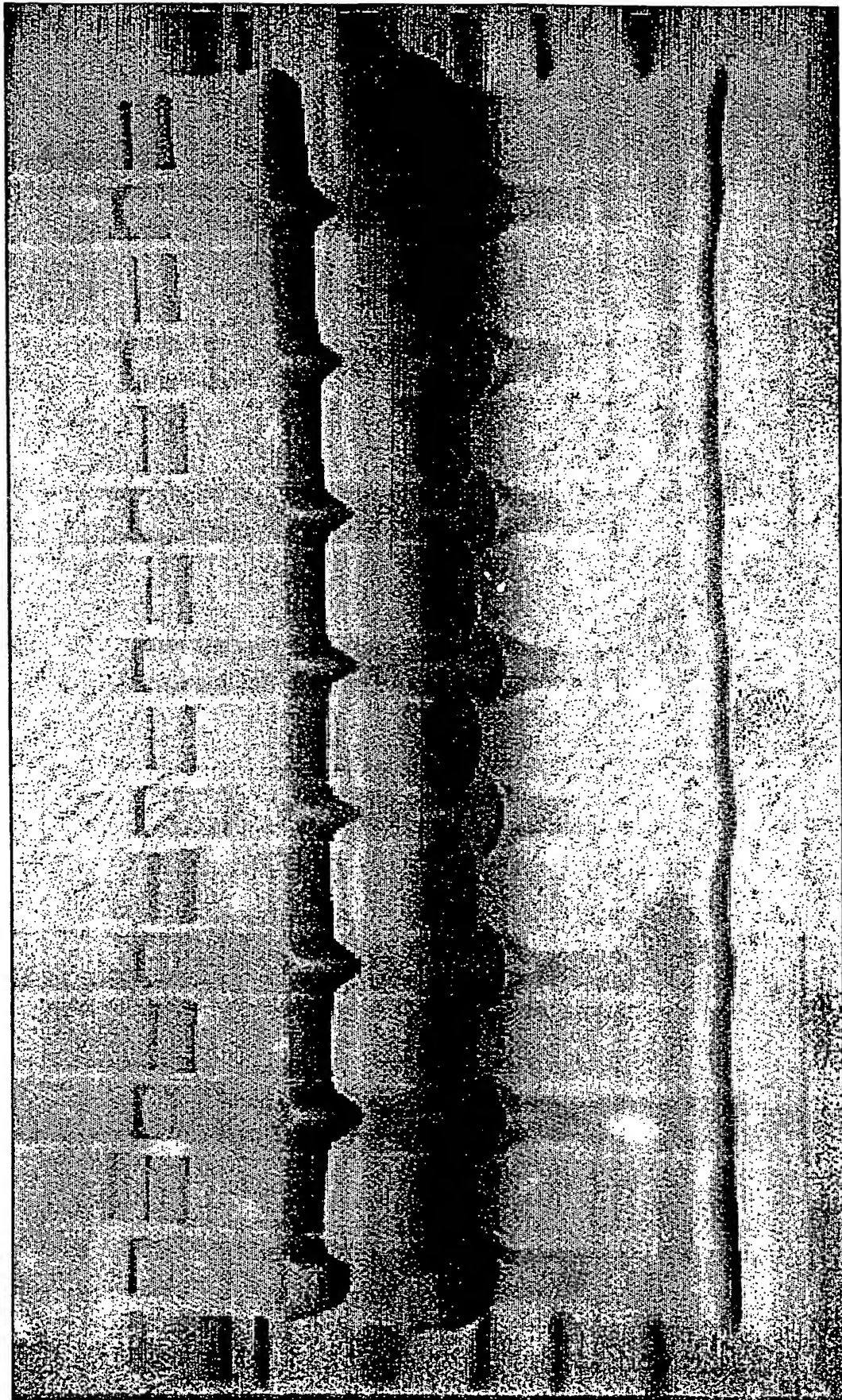


FIGURE 4

IS	Prom	SS	Gene of Interest	CS	Gene of Interest	polyA	IS
----	------	----	------------------	----	------------------	-------	----

**THIS PAGE BLANK (USPTO)**

FIGURE 5



**THIS PAGE BLANK (USPTO)**

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**